Web Controlling of Large Intranets

DISSERTATION
DER WIRTSCHAFTSWISSENSCHAFTLICHEN
FAKULTÄT
DER UNIVERSITÄT ZÜRICH

zur Erlangung der Würde
eines Doktors der Informatik

vorgelegt von

CHRISTOS KEFOS

von Widnau (SG) und Griechenland

genehmigt auf Antrag von

PROF. DR. LUTZ H. RICHTER
DR. REINHARD RIEDL

Juli 2005

Abstract

The Intranet has been set up in the enterprises as a medium of information. Even if it was seen as a technical feature of the IT department some years ago, a paradigm change has recently taken place during the last years. The Intranet is no longer settled into the IT department, but is now an integral component of the corporate communication. It is becoming more and more important because the organization members use the Intranet systematic for the spreading of information.

On higher demands of the company and his members towards the procurement and distribution of information, and also as the steady pressure to the effectiveness and efficiency increase, the Intranet is transforming from an information medium to a communication medium to a working instrument.

From the fact that more and more information is distributed through the Intranet, the organization members are increasingly confronted with the problem of flooding information: Documents cannot be found, information does not reach the suitable receivers or important instructions of the management are overlooked. As a consequence, discouragement and frustration spreads among the users. On the other hand, the company wants to make information available to all of its employees in order to reassure that knowledge is on the same level or to improve the effectiveness and raise the employees' contentment. In doing so, the company often does not know whether the action taken have led to success. The reality is often another where the company thinks that the employees are provided with the desired information, whereas the employees are disappointed because of the lack of information quality.

We have first conducted interviews to find out how the opinion of the enterprises towards the Intranet is, and also to find out how they use it to counteract against the flood of information and mismatch. Therefore, three signs which have validity for all enterprises can be chosen. First, the Intranet is recognized as a central tool of the corporate communication and is no longer considered as a mere toy of technology freaks. Second, the Intranet is regarded as an important information medium. Lastly, controlling the improvement of information quality is carried out by the least enterprises. Furthermore, the results from the interviews showed that the enterprises want to speed up the transformation of the Intranet. However, it is also clear that they are only at the beginning of the transformation path.

To raise the quality of information, we have developed a web controlling framework. This contains a web controlling process, which arranges the steps for the creation of the information space and a web perspective pattern, which generates different views of this space. Through this procedure and with the different views, the ability to form the information space is conducted.

As we have explained above, it is insufficient only to provide information and to believe that the needs of the users are satisfied with that. The information mismatches between employees and company can be caused either by excessive demand of the

employees with the exposure to the medium, or when the information is not present where the users logically expects it to be. In order to find out how the user moves in the information space, we have developed the generic filter pipeline. This filter pipeline derives the user behavior on the basis of web log file data. This user behavior reflects the real behavior of the information space users. In this way it is possible to indicate the discrepancy between the information offer of the company and the information search of the users.

On the basis of the results ascertained from above, the information space can be adapted to the user with the help of a derived Boehmian spiral, in order to achieve a continuous improvement of the needs and the surfing behavior of the user. However, since the information offer and the needs of the user change dynamically, the Boehmian spiral runs through regularly.

Zusammenfassung

Das Intranet hat sich mittlerweile in den Unternehmen als ein Informationsmedium etabliert. Wurde es noch vor ein paar Jahren als ein technisches Feature der IT Abteilungen angesehen, so hat in den letzten Jahren ein Paradigmenwechsel eingesetzt. Das Intranet wird nicht mehr an die IT Abteilung angesiedelt, sondern ist nunmehr ein integraler Bestandteil der Unternehmenskommunikation. Es gewinnt immer mehr an Bedeutung, denn die Organisationsmitglieder setzen das Intranet gezielt zur Informationsverbreitung ein.

Durch die immer höheren Anforderungen der Unternehmung und seiner Mitglieder an die Informationsbeschaffung und –verteilung also auch durch den stetigen Druck zur Effektivitäts- und Effezienzsteigerung, transformiert sich das Intranet vom Informationsmedium über das Kommunikationsmedium hin zum Arbeitsinstrument.

Dadurch, dass immer mehr Informationen über das Intranet verteilt werden, sehen sich die Organisationsmitglieder vermehrt mit dem Problem der Informationsüberflutung konfrontiert. Dokumente können nicht gefunden werden, Informationen gelangen nicht an die entsprechenden Empfänger oder wichtige Weisungen des Managements werden übersehen. Als Folge davon macht sich Demotivation und Frustration unter den Benutzern breit. Auf der anderen Seite möchte das Unternehmen seinen Mitarbeitern Informationen zur Verfügung stellen, um den Wissensstand aller auf die gleiche Ebene zu heben, um ihre Effektivität zu verbessern oder um die Mitarbeiterzufriedenheit zu erhöhen. Dabei weiss das Unternehmen vielfach nicht, ob die ergriffenen Massnahmen zum Erfolg geführt haben. Die Realität sieht dann meist so aus, dass das Unternehmen glaubt den Mitarbeitern die gewünschte Information bereitgestellt zu haben, diese aber enttäuscht sind über die mangelnde Informationsqualität der Unternehmung.

Wir haben Interviews durchgeführt, um erstens herauszufinden, wie die Haltung der Unternehmungen dem Intranet gegenüber ist, zweitens wie sie es benutzen, um der Informationsüberflutung und –mismatch entgegenzuwirken. Dabei können drei Merkmale herausgegriffen werden, die für alle Unternehmen Gültigkeit haben. Erstens, das Intranet wird als zentrales Werkzeug der Unternehmenskommunikation anerkannt und gilt nicht mehr als Spielzeug von Technikfreaks. Zweitens, das Intranet wird als ein wichtiges Informationsmedium angesehen. Und drittens, Controlling zur Verbesserung der Informationsqualität wird bei den wenigsten durchgeführt. Weiter ist aus den Interviews herausgekommen, dass die Unternehmungen die Transformation des Intranets weiter vorantreiben wollen. Es ist aber auch klar, dass sie sich erst am Anfang des Transformationspfades befinden.[1]

Um die Informationsqualität zu erhöhen, haben wir das Web Controlling Framework entwickelt. Dieser beinhaltet einen Web Controlling Prozess, der die

[1] See Technical Report concerning the outcomes of the interviews, [KeRi05].

Schritte zur Gestaltung des Informationsraumes festlegt und ein Web Perspektivenschema, das verschiedene Sichten auf diesen Raum generiert. Durch den Prozess und die verschiedenen Sichten soll der Informationsraum effektiv gestaltet werden können.

Wir wir oben festgestellt haben, genügt es nicht nur Informationen bereitzustellen und zu glauben, dass die Bedürfnisse der Benutzer damit befriedigt würden. Der Informationsmismatch zwischen den Mitarbeitenden und der Unternehmung kann entweder durch die Überforderung jener mit dem Umgang des Mediums sein oder die Information befindet sich nicht dort wo die Benutzer sie logisch vermuten. Um herauszufinden, wie sich der Benutzer im Informationsraum bewegt, haben wir die Generische Filterpipeline entwickelt. Diese Filterpipeline leitet das Benutzerverhalten anhand von Web Logfile Daten ab. Dieses Benutzerverhalten widerspiegelt das reale Verhalten der Informationsraumbenutzer. Dadurch ist es möglich die Diskrepanz zwischen dem Informationsangebot der Unternehmung und der Informationssuche der Benutzer aufzuzeigen.

Auf Basis der oben ermittelten Ergebnisse soll mit Hilfe einer abgeleiteten Böhmschen Spirale der Informationsraum durch eine kontinuierliche Verbesserung an die Bedürfnisse und das Surfverhalten der Benutzer angepasst werden. Da sich aber das Informationsangebot und die Benutzerbedürfnisse dynamisch ändern, wird die Böhmsche Spirale regelmässig durchlaufen.

Acknowledgement

My first and foremost thanks go to my thesis advisors, Prof. Dr. Lutz Richter and Dr. Reinhard Riedl of the Department of Informatics (IFI), for giving me the opportunity to write the dissertation. Especially, I would like to thank both for the close supervision and the countless scientific discussions, which enabled me to finalize the thesis.

I would like to thank the staff of Department of Informatics at University of Zurich for both providing the required infrastructure to make my work at the Department possible and helping me with administrative affairs.

Furthermore, I would also like to thank my interview partners for both taking the time to conduct the interviews and giving me the opportunity to get an insight into the practical application of the Intranet.

Last but not least, I would like to thank my wife Evdokia and my daughter Aglaia for supporting me, taking me off the duties of daily life and being appreciative of my frequent absence from the family.

To my parents

Contents

List of Figures

List of Tables

1. Introduction

1.1 Problem and Motivation

The initial point of this thesis was a project conducted at a large bank in the Swiss market place. It was supported by the Swiss National Fund (SNF) of the Swiss government. The project addressed the topics of *large Intranets*, especially the problems which occur during its use and growth. The bank was interested in the research of those problems because its Intranet grew continually and the use became increasingly inefficient. This inefficiency was reflected in various ways: In the rise of the response time of the requests, in the decline of the hit precision of search requests, in the decline of the use, in the impossibility to find needed information or documents, and in the decline of the usage of provided tools to facilitate the Intranet use. As a result from inefficiency and refusal of using these tools, other problems such as de-motivation and annoyance of the user arose. A common reaction from the employees was that one could never find a document on the Intranet unless one has the exact URL of the document.[2]
The results were obtained from our hands-on experience of the Intranet and a loose poll of some employees of the bank. The results from our research are listed below and will be explained subsequently.

- Many sources of information
- Heterogeneous culture of information
- Insufficient quality of information
- Insufficient quality of service
- Dominance of the pull principle

Sources of information:
The bank's Intranet had many sources of information which confused and also asked too much from the users. In the following we list the sources and give a short explanation of their purpose. There was among others a *DocWeb*, a part of the Intranet which contains only documents written in .doc, .pdf or .ps formats. W*ho's who* is a web-based request service to retrieve the employee's profile like business phone number, email address, the department they are working for, position, and job description. *BWsearch* stands for Bank Web search and is a search service to retrieve miscellaneous information from the Intranet. *Informia* is a meta-search engine which activates all services. It is defined in its domain, to conduct the requested search by forwarding the request to those services, collecting the results afterwards and displaying them to the user. *WebCC* is the information space of the Web Competence Center where users can find miscellaneous information on IT technology, projects, company news, personal websites of the employees and etc. In addition to the variety of the sources another problem was the huge amount of data. The bank had about 100 web servers, about 200'000 web pages and different entry points to the Intranet. It is needless to mention that the users were asked too much.

[2] See [LuRi00] for further insights of this problem.

Heterogeneous culture of information:
The heterogeneous culture arose when the bank merged with two other banks into one. The IT departments recognized that they had different cultures and perceptions of information representation. As a result there were different representations of the websites on the Intranet. Different representation also means different web structure, different navigation paradigm and so forth. With that consequence that by changing the sites users were irritated and they were obliged to find their way again. In the post-merger phase, decisions were taken on a high level to integrate the three bank and their businesses smoothly. Unfortunately, the strategical and political decisions paralyzed or stopped existing and influenced future projects. A main point to consider after the merger was also the fact that different company cultures clashed. This brought about consequences in the behavior of the employees, that a tendency to deny access to information to employees from the other former banks arose.

Insufficient quality of information:
Thereunder, we understand the three subjects: Garbage collection of invalid information, quality management of the information publishing process, and the high dynamic of the environment. First, there is no garbage collection carried out on the Intranet with the result that the information graveyard becomes bigger and bigger. Old or invalid information is left carelessly on the Intranet. Furthermore, there also exists an insufficient update policy, which leads to the fact that information is available in different versions and places. In the end, the user does not know which information is accurate and which is not. This becomes a big issue if it concerns directives from the top management level concerning non-disclosure policy of the company, vacation policy, travel expenses policy, off-the-job training policy and many others. If old information is omitted to be remove then it can create a climate of uncertainty within the company and old directives will mistakenly be taken as accurate and executed.

The second subject deals with quality management resp. with the absence of it. The company has no justification enforcement concerning the publication of web pages on the Intranet. At first glance, this might look unimportant. However considering the fact that the company has about 70’000 employees and each one is allowed to create web pages and put on the Intranet, the impacts can be tremendous. The second point to be mentioned is that the bank has no policy in place, which defines necessary quality standards describing the structure, layout, versioning, updating of web pages. Quality guarantees and controls are missing and web pages do not have to go through a quality acceptance process before being published. A further point focus on the absence of a signature principle and control. Due to the missing signature, it is seldomly obvious who is responsible for the content of a web page. As mentioned above, the impact of wrong information on a web page is negligible if it is about new technology, project progress, employee information and of likewise. Yet if it is a matter of company directives or important company information then it is important for the employee to see that the content is correct and trustable.

The third subject in which we analyzed is the high dynamic of the environment. Due to the merger of the financial institutes, the restructuring of the websites was very dynamic. Old websites resp. web pages were removed, changed, or updated and new websites resp. web pages were created. The whole process occurred at a very high

pace. The merger also made employees to leave the bank with the result that the owners of the particular websites or web pages resp. were lost. As a result, the information graveyard augmented once more.

Insufficient quality of service:
We looked at the quality of service from the user's and the administrator's point of view. On the user's site, the quality of service affects on the one hand the inadequate support of the users for the Intranet usage. On the other hand, the availability of some services was limited. An example is the who's who service at the start-up phase. On the administrator's site, the quality of service affects the internal firewalls, which still existed after the merger as well as the insufficient network monitoring infrastructure.

Dominance of the pull principle:
This means that every employee should be self responsible in order to get the necessary information that he or she needs. There is no notification for the employees that a new directive is put on the Intranet. They have to periodically check the Intranet for new ones. The drawback here is that old entries are not removed. This situation is a potential source to cause uncertainty.
In our opinion the end-users are neither treated as customers nor are they educated as professional Intranet users. It is a trial and error approach.

During the research we saw that one of the reasons for the difficulties with the Intranet interaction lie in the missing Intranet culture. People were neither used nor trained to deal efficiently with this new communication media. Another point was that the bank did not recognize the economic relevance of the Intranet.
A typical method of resolution was to build new tools and services. An example is the meta-search engine Informia. The search engine activates several services simultaneously, neglecting the fact that by enlarging the search space the precision declines and unnecessary work load is generated.

Starting from the research results, we identified the problem of the mismatch between the user's need and the provider's offer of information. Based on that fact, we derived some questions which we intended to investigate further. Some of the interesting questions in our perspective were:

1. Which interests of the users can be extracted from the website
2. How should the design and the supply of a website be adjusted in order to match the preferences of the user
3. What is the actually surf behavior of the user and how are the available tools used

To answer the above questions, we assume that we need to understand the real behavior of the users in a particular environment. We also assume that an image of user behavior is stored in the web server log files. We introduce the web controlling framework to find out the user's needs and behavior, to evaluate the information space implementation and to identify the tools used.

1.2 Objectives of the Thesis

As we have seen from the previous chapter *Problem and Motivation,* there is a mismatch in the perception between the company and the organization members, regarding the information needs of the latter and the information offer of the former. This discrepancy first comes from the inadequate design of the information space, second from the unattainability of information by the users, and third from the company's lack of knowledge about user's information needs. Another difficulty is that the information space is dynamic: First, because documents are added and removed from it on a regular basis, and second, because with time the users change their behaviors. Latter is caused by the learning effect of the users. A few years ago, users had the additional problem that they did not know which source the information came from, who is in charge of it or if it was still valid. Enterprises responded to that by using content management systems (CMS) increasingly. The CMS is used to publish content. It allows providing information with a creation and an expiry date. Furthermore, on each web page, the person in charge is denoted. This approach is rather sender oriented. The focus here lies on the source of information. Enterprises additionally provide search engines for information search in their Intranets.

In our thesis, we reverse the focus and propagate a user-oriented focus. We state that in order to reduce perception differences, it is important to know how users behave in the information space. We further state that the user behavior is stored in the log files. The objective of the thesis is to provide an integral approach which on the one hand enables the design of the information space, and on the other hand, takes into account user's behavior. Moreover, user behavior must influence the information space design since we demand user orientation. The information space is used in the context of web based Intranet.

We will accomplish that by introducing the web controlling framework. This framework enables the controlling process which connects the system design with the user behavior. It is subdivided into two parts: The Inside-out and the Outside-in view. Former is the view of the company on the information space. It defines the necessary actions to be implemented in order to form the information space. Latter looks at the users in the information space and extracts their behavior from collected data. As mentioned above, these data are the web log files.

We further introduce the spiral model. The objective of the spiral model is to support continuous improvement of the information space based on user behavior. Continuous improvement is warranted by introducing a cyclical process.

In order to analyze user behavior, we implement the generic filter pipeline. This pipeline analysis collected data such as log files. Since the architecture is open, different functionalities can be plugged to the pipeline in order to process the data.

Our research methods are interviews with experts and users, formal modeling with elementary algebraic and stochastic methods, and the analysis of web log files, which were evaluated both manually and stochastically.
The research began with a cojoint project with a Swiss bank, where we conducted a study about the intranet of this bank.

Our methods are based on research collaboration with the above mentioned Swiss bank and interview with several enterprises. Former provided an insight into the problems a company with a large Intranet faces. Latter made it possible to see how enterprises nowadays use the intranet. From the interview, we gathered the necessary requirements for our information space.

The main result is the controlling framework which evaluated the technical feasibility by means of the generic filter pipeline.

1.3 Organization of the Thesis

The rest of this thesis is organized as follows. Figure 1 gives a graphical overview of the thesis' chapters and their relation to each other.

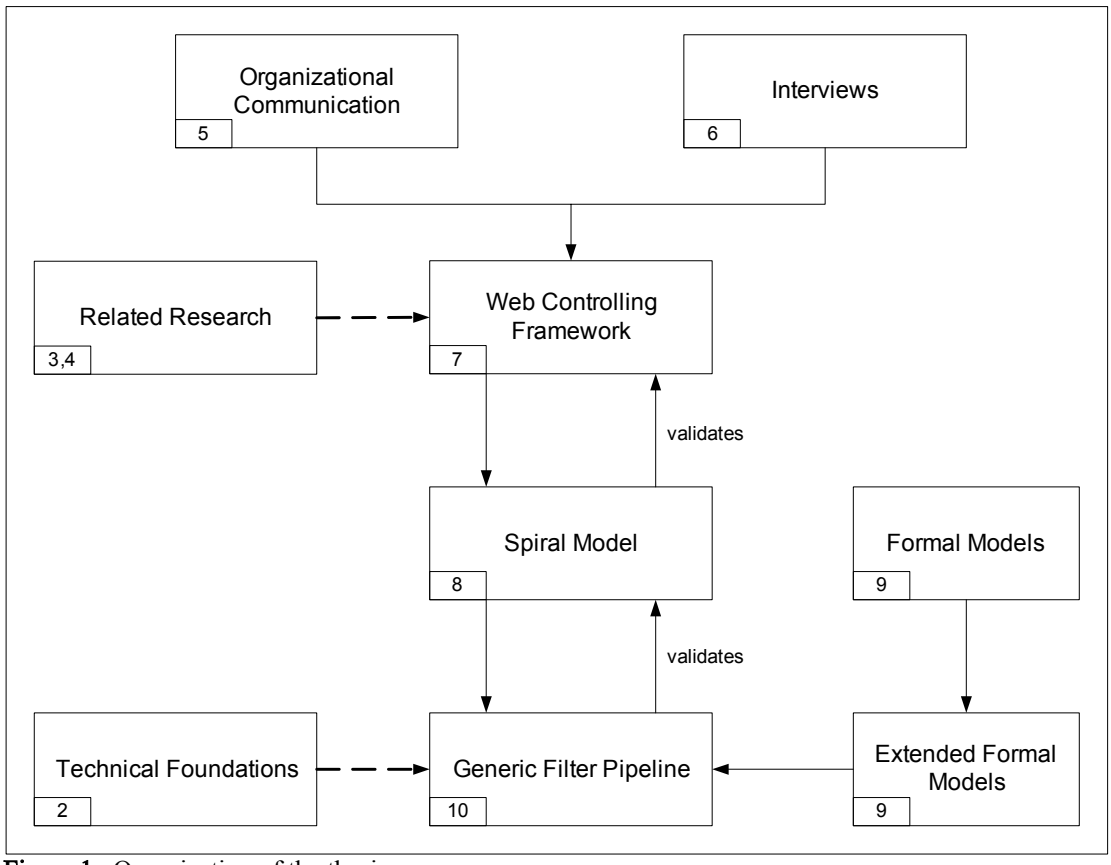


Figure 1 Organization of the thesis

Chapter 2 introduces the research area of web mining and presents the technical foundation for its analysis. It further discusses how pattern discovery and analysis is conducted. This chapter lays the basis for the implementation of the generic filter pipeline.

Chapter 3 discusses related research areas on where and how web mining is applied to and gives as example insights to web personalization, web organization, virtual communities, costumer profiles, eMarketing and benchmarks.

Chapter 4 discusses the thesis' research topic web controlling and gives a definition and an introduction of it. Further, it states which the basic principles are it builds on.

Chapter 5 discusses organizational communication. Since the Intranet is acting as an information channel, it is essential to look at organizational communication and its goals and tasks in more detail. Further, the basic statements given in corporate communication are tried to be confirmed or confuted based on the interviews.

Chapter 6 summarized the interviews. The main topics of enterprises' goals and purposes to the usage of the Intranet are presented.

Chapter 7 introduces the Web Controlling Framework. It discusses its purpose, gives an introduction of the controlling process, the perspective schema, the perspective framework and explains its usage.

Chapter 8 discusses the spiral model. It is based on the Boehmian spiral and expanded further to fulfill our needs. The spiral model assures a continuous improvement of the information space design and usage.

Chapter 9 introduces the technical view of an information space. Therefore, it presents several models, its relations and its characteristics.

Chapter 10 introduces the generic filter pipeline. The pipeline is used to analyze the web logfiles and to derive further actions for the information space usage.

Chapter 11 validates the requirement. It outlines if each requirement presented in chapter 7 has been satisfied.

Chapter 12 summarized the contributions of the thesis. It gives an outlook about the development of the Intranet as a communication medium in enterprises and discussed further research in this topic.

2. Technical Foundations

This chapter introduces the concept of web mining. It gives a definition of it and of current web mining classifications. Afterwards, the focus is directed to web usage mining, which deals with the analysis of user behavior. It explains its useful purposes and which role it plays for the thesis. Finally, pattern discovery and pattern analysis are introduced. This chapter aims to lay the technical foundation for the formal models in chapter 9 and the implementation of the generic filter pipeline in chapter 10.

2.1 Introduction to Technical Foundations

In this section, we give an introduction to the technical foundations of our work. These are lain in the research area of web mining, which deals with the discovery of knowledge and patterns from web data. These data can be collected at the server-side, client-side, proxy servers or obtained from an organization's database. Srivastava et al. classify web data into the following types [SrCo00]:

- Content: The real data in the Web pages, i.e. the data the Web page was designed to convey to the users. This usually consists of, but are not limited to, text and graphics.

- Structure: Data which describe the organization of the content. Intra-page structure information includes the arrangement of various HTML or XML tags within a given page. These can be presented as a tree structure, where the <html> tag becomes the root of the tree. The principal kind of inter-page structure information is hyperlink connecting one page to another.

- Usage: Data that describes the pattern of usage of Web pages, such as IP addresses, page references, and the data and time of accesses.

- User Profile: Data that provides information about users of a Web site. A user profile contains demographic information, such as name, age, education, country, for each user of a web site as well as information about users' interests and preferences. Such information is aquired through registration forms or questionnaires, or can be inferred by analyzing Web usage logs.

Web mining describes the universal application of methods for common data mining to data for the web. This contains the analysis of page content (web content mining), of page structure (web structure mining) and of user behavior (web usage mining) [HiMe02]:

- *Web Content Mining* deals with the analysis of the content of web pages. The goal is the facilitation of the information search in the web. Assignments are

classification and grouping of online-documents or the locating of documents according to particular search terms.

- *Web Structure Mining* analyses the arrangement of several elements within a web page (intra-page structure information) as well as the arrangement of several web pages to each other (inter-page structure information). Notably, the references from one page to the other by means of hyperlinks.

- *Web Usage Mining* deals with the behavior of web users. Data mining methods are applied to web server log files to identify user patterns and interests and discover knowledge.

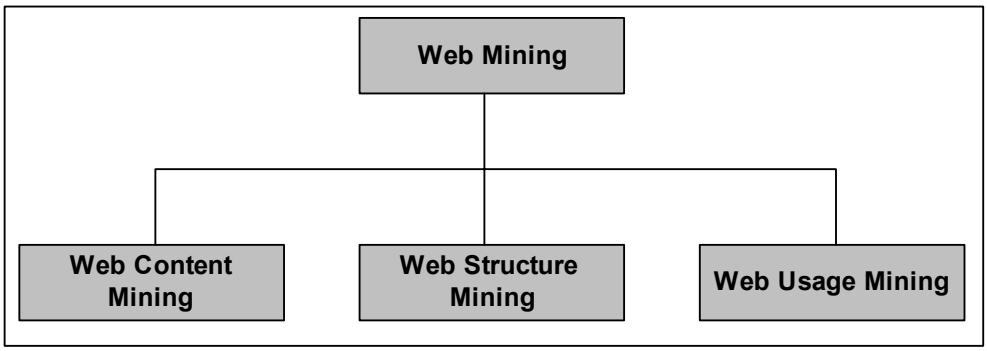


Figure 2 Source [HiMe02], Web Mining Classification.

Figure 2 shows the web mining classification tree. If the analysis of user behavior is limited to web server log files, then we talk about Web Log Mining. There are more data sources used then we talk about Integrated Web Usage Mining. The additional data sources include user registration data, referrer logs which contain information about the referring pages for each page reference, and survey data gathered by various scripts [CoMo99]. Figure 3 shows the conceptual classification of Web Usage Mining [HiMe02].

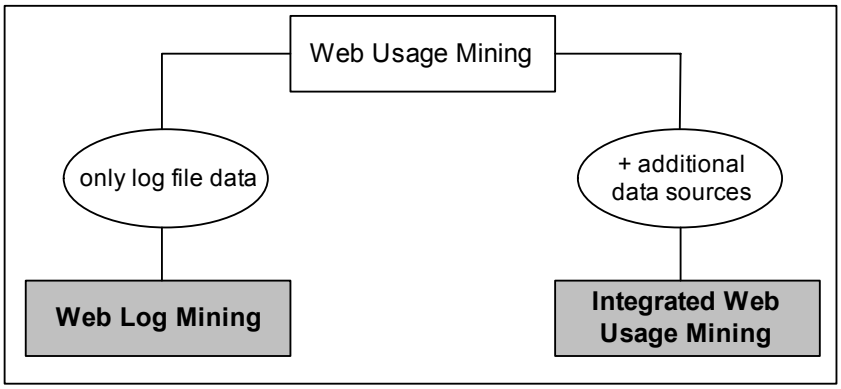


Figure 3 Source [HiMe02], Types of Web Usage Mining.

In the thesis, we are going to concentrate on Web Usage Mining.

In the following section, we will describe the mining process. Before we concentrate on our primary focus, which is the discovery of knowledge and user behaviour, we need to preprocess the web log file data. Hippner et. al [HiMe02] introduce the transaction of the mining process as described in figure 4.

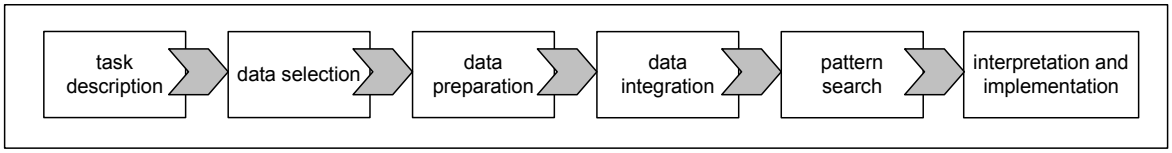


Figure 4 Source [HiMe02], Web Mining Analysis Transaction

The necessary data are extracted from the log file depending on the tasks. The subsequent data preparation phase is subdivided into the two steps; data cleaning and user and session identification. These phases are important for the next phases, but also consume the most time within the process [HiMe02].
Mobasher et al. [MoCo00] describe a similar approach of the mining process (see figure 5).

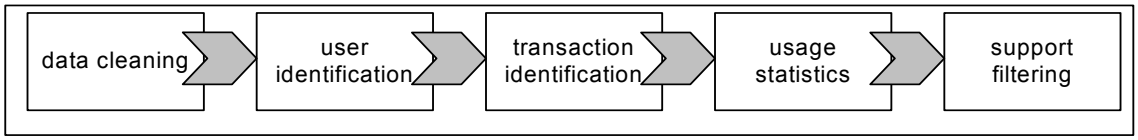


Figure 5 Source [MoCo00], Web Mining Process

A more detailed mining process gives Cooley et al. [CoMo97] in figure 6.

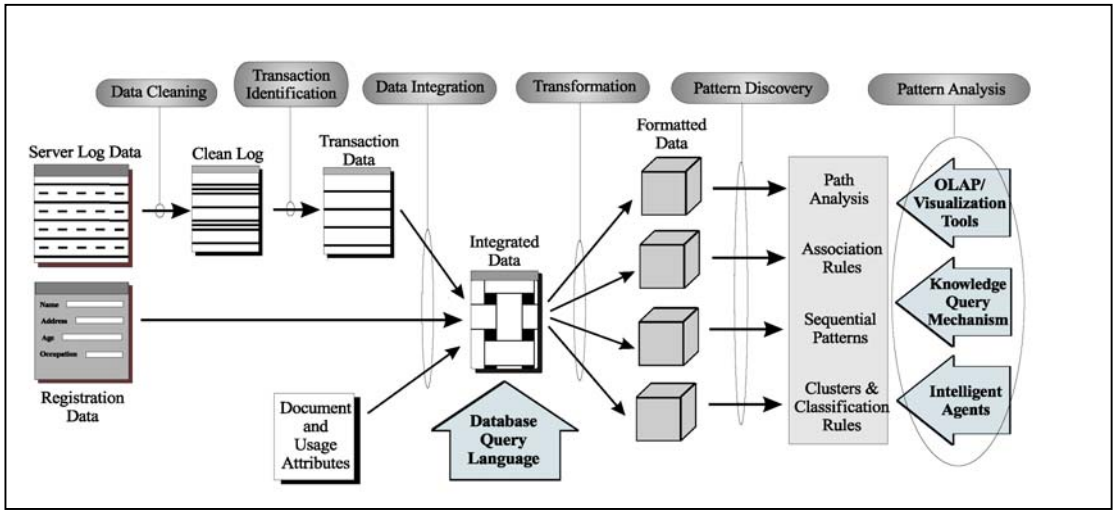


Figure 6 Source [CoMo97], General Architecture for Web Usage Mining

Our generic filter pipeline[3] follows the above described approaches insofar that we also clean the data at the beginning. I.e. we remove all .gif, .jpeg, .mpeg files as well as all spiders and robots. Afterwards, we extract the important data like IP address, Time, Request Object and History and generate user sessions. It must be said that we

[3] See chapter 10, Generic Filter Pipeline, p. 109 et seq.

do not use additional data sources like email, user registration data, and others that are similar.

2.2 Data Sources

As we have stated, the data collection and analysis are based on web server log files. Additional data which can be used are cookies, user data from register forms, demographic data, campaign information (banner, email), transaction data (buy, order) [HiMe02].

2.3 Preprocessing Tasks

The preprocessing tasks are divided by the experts into the same three subtasks. The first subtask is data cleaning. The raw log file contains many information which are irrelevant for web log analysis. These information can be removed by checking the URL suffix names. All entries with suffixes such as map, jpeg, JPEG, gif, GIF can be removed. Entries which status code indicates an error in the transmission should be removed, too. Furthermore, robots and spiders must be removed as well because they do not give any useful information on the subsequent analysis process.

The second and the most difficult task is user identification. Unless a client side tracking mechanism, e.g. cookies, is used, only the IP address, agent and server side clickstream are available to identify users and sessions. Some of the problems are [SrCo00]:

- Single IP address/Multiple Server Sessions: Internet service providers (ISPs) typically have a pool of proxy servers that users access the Web through. A single proxy server may have several users accessing a Web site, potentially over the same time period.

- Multiple IP address/Single Server Session: Some ISPs or privacy tools randomly assign each request form a user to one of several IP addresses. In this case, a single server session can have multiple IP addresses.

- Multiple IP addresses/Single User: A user that accesses the Web from different machines will have a different IP address from session to session. This makes traking repeated visits from the same user difficult.

- Multiple Agent/Single User: Again, a user that uses more than one browser, even on the same machine, will appear as several different users.

An approach to identify different users within the same IP address is to join them with the browser or by using the operating system to infer thereof the different users. Even

the use of the cookies can only deter the computer used. If multiple users use one computer, than only a registration helps to determine each user separate.
After having irrelevant data removed and identified the potential users, we focus on the third task, the session identification. A session is a set of requests from the same user with a maximum time difference of 30 minutes (as described in [SrCo00]) between subsequent requests. It is an ordered sequence of pages requests from that user. The problem one faces is to identify the log entries which belong together. If no session id is assigned then a good indicator, as mentioned in [SrCo00], is the time difference between subsequent requests.

2.4 Pattern Discovery

Having finished the preprocessing task, we continue with the pattern discovery task. There have been applied methods and algorithms from several fields such as statistics, data mining, machine learning and pattern recognition [SrCo00]. In the following, we are going to describe the mining activities that have been applied in the web context. The summary is extracted from Srivastava et al. [SrCo00], Mobasher et al. [MoCo00] and Riedl [Ried01b].

1. Statistical Analysis
   Statistical techniques are the most common method to extract knowledge from web data. By analyzing the data, one can perform descriptive statistical analyses such as frequency, mean, median, most frequently accessed pages, average view time, and variables such as pages, session length etc. The generic filter pipeline contains a component which generates some of those statistics.

2. Association Rules
   Association rule algorithms try to discover all associations and correlations among data items, where the presence of one set of items in a transaction implies the presence of other items [MoCo00]. An example for correlations between web pages is the following:
   - 70% of users who accessed the page /company/info/, also accessed the page /company/info/info1.htm (s. figure 7)
   - 30% of users who accessed the page /company/info/info1.htm, also accessed /company/info/info2.htm

   Association rules help to organize the company's information space clearer, more user-friendly and more structured. For example, the company is able to discover that 85% of the users accessing /company/info and /company/info/info1.htm also accessed /company/info/info2.htm, but only 25% of those who accessed /company/info also accessed /company/info/info2.htm.

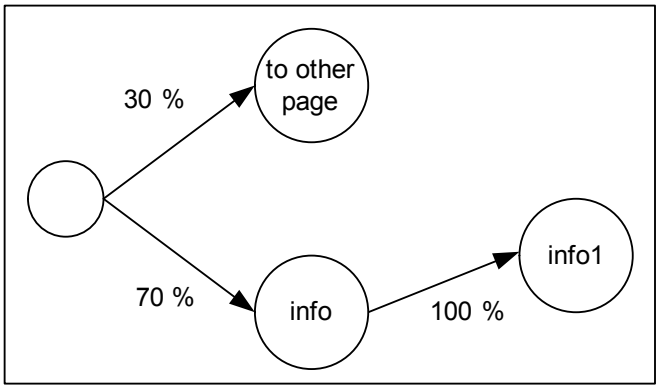


Figure 7 Gives a visual interpretation of the rule above.

We might infer that some information on info1 leads users to info2 and therefore, these information should be moved to a higher level, e.g. /company/info, to increase the access to figure 7.

3. Sequential Patterns and Time Sequence
   A prerequisite is that the timestamp of the sessions is explicitly available. The goal is to find inter-session patterns such that the presence of a set of log entries is followed by another log entry in the timestamp ordered session set. An example is:

   - 30% of the users who visited /company/info/info1.htm has done a search on Google within the past week on keywords k_1 and k_2 (s. figure 8).

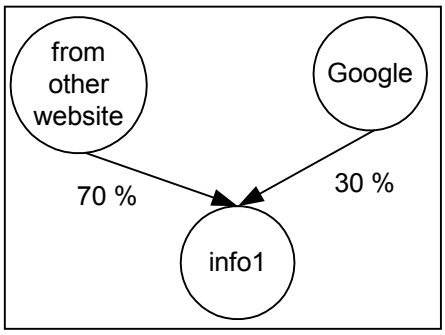


Figure 8 Interpretation of sequential pattern and time sequenc.

Similar time sequences is another interesting dependency. For example, we may be interested in finding common characteristics of all clients who visited a particular file within a certain time period. Or, we may be interested in a time interval in which a particular file is most frequently accessed [CoMo97].

4. Clustering
   Clustering is a technique to group together a set of objects having similar characteristics. Clustering of users discovers groups with similar behaviour. Such information can be used to provide personalized web content to the users [MoCo00].

5. Classification
   Classification is a technique where one pre-defines groups with particular attributes. This is a kind of profile of a particular group. The profile can then be used to classify new sessions or log entries. The classification technique can be applied, for example, to users by creating profiles for different users. The relation that is discovered is as follows [MoCo00]:
   - Users who access /company/tech/ tend to be from the IT department.
   - User who accessed /company/misc/pensionplan.htm in the 60-65 age group.

6. Dependency Modeling
   Dependency modeling is another pattern discovery technique for web mining. The goal here is to develop a model capable of representing significant dependencies among the various variables in the web. As an example, the company may be interested to build a model representing the different stages a visitor undergoes while shopping in an online store based on the actions chosen, i.e. from a casual visitor to a potential customer [SrCo00].

7. Frequent Item Set
   Association rules find groups of web log entries which occur frequently together. These groups of log entries are referred to as frequent item sets.

2.5 Pattern Analysis

In order to extract knowledge from the patterns we have discovered, we need to analyze them as a last step. Thus, pattern analysis is the last step in the web mining process. The motivation behind it is to extract meaningful knowledge. In a way, that can be applied to help reach the overall goals of the company.
The exact analysis methodology is usually governed by the application for which web mining is done. The most common form of pattern analysis consists of a knowledge query mechanism, such as SQL. Another method is to load usage data into a data cube, in order to perform OLAP operations [SrCo00].
There are different tools which exist to perform pattern analysis as well. Given in a short list, these are: Surfstats Log Analyzer, FastStats, Azure Web Log Analyzer, WebTrends, eIQ Web Analyzer, WebLog, Webalizer, etc.
In this section, we would like to give a brief evaluation summary of the above Web Mining Tools. In general, all of them have the same features. These are domain, name, date, hits, page views, sessions, bandwidth, browser, and operating system. In addition to the general information, the tools provide also more detailed information such as hits, total data transferred, total visiting users, average hits per user, average users per day, average data transferred per user, and average hits per day. Some tools also provide information about top entry page, top exit pages, navigation path and visit duration. This gives a rough idea of which path the user has taken through the site and how much time he or she spends to look at the pages. All tools support the

standard Common Logfile Format for server log as well as some variations of the Combined Logfile Format. Common and standard to all, is the graphical user interface to display the results. The tools also support multiple languages.

2.6 Conclusion

This chapter showed the state-of-the-art implementations of web mining. It shows that the preprocessing tasks such as data gathering, data preparation and data integration are the same for any analysis tool. Our generic filter pipeline is no exception in this matter and therefore follows the same tasks. Its difference from the pattern analysis tools described above is that our generic filter pipeline has an open architecture. This means that in addition to the existing build-in functionalities, it is possible to add new functionality. This enhances the flexibility and makes our analysis tool more adaptive to the analyst's needs.

3. Related Research Areas

This chapter introduces the various areas of web usage mining application. Web usage mining, as already defined, deals with user behaviour which is of importance for us. Following is a short introduction to some of those areas that are given and outlined on where and how web usage mining is applied:

- Web Personalization
- Website Organization
- Customer Profiles
- Virtual Communities
- eMarketing
- Benchmarks

3.1 Web Personalization

Web personalization is the process of customizing the web site to the needs of specific users, taking into account the knowledge acquired from the analysis of the user's behavior in relation to the user profile data, web structure and web content [EiVa03]. Thus, personalization is concerned with what to present to the user, how to present it and how can the user be guided through the site according to his preferences.
The personalization process is mainly divided into the three phases; profiling, matching and channeling. The aim of the profiling phase is to generate a comprehensive picture of the user. This is done by generating user profiles. These profiles are, in turn, created by integrating relevant internal and external data. Internal data are log files, cookies, etc. External data are age, gender, domicile, preferences, etc. The matching phase deals with the merge of user profiles with the particular content. The goal is to present the appropriate content to the user. The channeling addresses the question of the appropriate distribution of the content to the user, such as ordinary mail, Internet [Gent02].
Mobasher et al. [MoCo00] present an automatic personalization concept based on web usage mining. The purpose of the system is to provide real time recommendations of links to users based on their current navigational activities. These recommendations are derived from an offline preprocess and from an online process task. The overall process is divided into two components. The offline component is comprised of the data preparation tasks[4] and the online component is comprised of a recommendation engine and the web server.
The preprocessing tasks are batch processes. Starting from the raw of web server log files and by going through the entire process of web mining, it results in a user transaction file. Applying the web mining techniques, such as association rules and clustering, to the user transaction file, results in URL clusters and frequent item sets. The clusters and item sets are used by the online component to provide dynamic

[4] See chapter 2.3, p. 10.

recommendations. The online component is the recommendation engine of the web personalization system. The task of the engine is to generate a recommendation set for the current session. This set is based on the usage pattern, derived as described above, and consists of links to web pages which the user might want to visit. These recommended links are then added to the last page in the session, accessed by the user before that page is sent to him [MoCo00].
Eirinaki et al. [EiVa03] present the architecture necessary to conduct web personalization using web mining. The process of usage-based web personalization consists of five modules [EiVa03]:

- User profiling: User profiling is the process of gathering information specific to each user. A user profile includes demographic information, interests and navigational behavior of the user.
- Log analysis and web usage mining: The information in the web server log files is processed by applying data mining techniques. This process can be regarded as part of the user profiling process.
- Content management: The content of a web site is classified in semantic categories in order to make information retrieval and presentation easier for the users.
- Web site publishing: The content is presented through a web publishing mechanism to the user in order to guarantee a uniform way of data representation.
- Information acquisition and searching: Information is not only provided from the local web server. It can be also gathered from various web sources and therefore, the process of searching must be supported by relevant searching and ranking techniques.

The interaction of the above described modules result in the web site personalization for the particular user. Eirinaki et al. addresses the difference of personalization and customization. The latter is a static adjustment of user's preferences regarding its structure and presentation. The former is a dynamic adjustment of structure and content. The concept of personalization is also applied for our information space. What we have to stress here is the fact that unlike web personalization, the adaptation is not made individually for each user. The adaptation is made against the background of the efficiency of usage.

Meyer et al. ([MeWe01]) present a process model and tool for real time web site personalization with the help of web mining algorithms. In addition to the web server log files, they also use additional data. These data come from application server logs, for dynamic page generation or database requests, and from Java applets log files integrated into the web pages.
The process consists of two closed loops, the analytical and the real-time loop. The analytical loop is responsible for the preprocessing and analysis of the various log files. The preprocessed data is written into a database. Out of this database, a web data mart is generated and analyzed with OLAP and web mining techniques. The result of this process is the generation of rule types which describe typical user profiles. Since

the loop is closed, it allows the permanent update and generation of new rule types. The real-time loop is responsible for the personalization of the web site in real time. Therefore, it takes the generated rule types and applies them to the user, who is surfing on the web site. The real-time loop classifies the current user into a pre-determined user profile and personalizes the web site on real time with this appropriate user profile content. Through the use of additional data, the authors overcome the problem of session and user identification and therefore generate accurate user profiles. The process is limited to user profile identification and the mapping of current users to a profile. The concept of web controlling[5] is in this approach too narrowly interpreted.

Nakayama et al. ([NaKa00]) introduce a concept to identify gaps between the expectations of web site suppliers and the users behavior. The concept is based on a vector space model. Their assumption is that conceptually related pages should co-occur in visits if they are well designed. Thus, conceptually related pages that do not resp. rarely co-occur are not well designed and can be improved. To achieve this, there they first measure the inter-page conceptual relevance. A page is represented as a vector weighted with its word frequency. The relevance is measured for each page pair with a cosine similarity formula. Second, they also need to measure the inter-page access co-occurrence. This is done by generating a vector of IP addresses weighted with their frequency. The access co-occurrence is measured with a cosine similarity formula as well. Finally, they generate a plot of the inter-page conceptual relevance versus the access co-occurrence. From the result, one can derive two design improvements. One is a hyperlink topological improvement and the other is a page layout improvement. The authors built a system that gives quantitative suggestions on the redesign of page layout. With the suggestions given by the system, designers can improve the web site following those steps. Also, this approach is valuable to be applied to the information space. This approach is valuable for finding gaps between the two user groups. A drawback for our purpose is that the system is reactive and not progressive which means that the user shapes the form of the site and not vice versa. Over time, the site adapts to the user's behavior. It is not possible with this concept to change the behavior of the user.

3.2 Web Organization

Srikant et al. [SrYa01] introduce a method to organize website according to the visitor's expectations. This method is based on the assumption that the visitor will backtrack on the website if he does not find the information where he or she has expected. The point from where he backtracks is the expected location and the web page where the information lies it the target page. If an expected page has a significant number of hits, then the idea is to add a link between that page and the target page. To find such expected locations, the authors developed a novel algorithm which works like following. From early on, it has to be mentioned that web browsers

[5] See chapter 4, p. 21, for the definition of web controlling.

have a caching function which means that if the function is enabled the browser will not request the same page again, but will show the cached one. It is first disabled, then it will request the same page again from the web server. Why is that important? The data used by the algorithm are the requests in the web server log files. Thus, identifying the expected location with disabled cache means that the previous and next page in the web log file entry is the same. On enabled cache the expected location can be found based on the assumption that if there is no link between pages p_1 and p_2 the user must have hit the "back" button in the browser to go from p_1 to p_2. Thus, to find expected locations, the algorithm must first check if there is a link between two consecutive pages. As we mentioned above, identifying expected locations can help to organize the web site more efficiently and meet the visitor's expectation. The results then can be handed on to a web designer or an administrator, who will enhance the site accordingly. Web Organization is not part of our thesis because it deals with the retrieval of already available information. We look more at the fact whether we have the right information on the information space and how to determine it.

3.3 Virtual Communities

Savaresi et al. [SaGa03] analyse a large virtual community. The analysis is based on two different data sets. The first data set is the user's database. Each user has to register to the virtual community and is given a unique nickname. The second data set is the web server log file of the virtual community. In their analysis, Srivaresi et. al establish a relationship between the user's database and the web log file of the community. They merge the two data sets to find relations between users DB and user web navigation behaviour of the virtual community. The approach of this task is as described in [SaGa03]:

- The users' DB has been analysed and clustered into a small number (12) of clusters. Each class represents a prototype of user.
- The log file of the web server has been first sessionized and then analysed and clustered into eight clusters. Each cluster represents a navigation behaviour.
- Because of the dimensional reduction of the two data sets, it is possible to find an association map between users and navigation sessions. This can be done because the registered users have a nickname, stored in a cookie. This allows the linking between users and log file entries.

For a better understanding of how the analysis is conducted, we refer to [SaGa03]. The information which can be drawn from the associations are, for example as follows: Males seems to be related to long and various sessions, females seem to be primarily interested to sessions with forum or chat. For further associations on this example, we refer to [SaGa03]. The fact of user registration makes the problem of user identification inexistent.[6] We will not touch the topic of virtual communities in

[6] See chapter 2.3, p.10.

our thesis, although, virtual communities can be integrated into the information space to facilitate the communication between the users and increase the intellectual capital[7] of the company.

3.4 Customer Profiles

Adomavicius et al. [AdTu01] describe their process of building user profiles using data mining. In order to do that, they developed the 1:1Pro (One-to-one Profiling) system. The system is based on the concept described subsequent. They state that the user profile is composed of two parts. One part contains facts about the user, and the other contains transactional data. Facts are data about the user like name, address, gender, employee number, and birth date. Transactional data describe users' actions like navigation path, documents read, items downloaded, and so forth. After having identified these data, the system generates rules. These are derived from the above mentioned data and describe users behaviour. To give an example, such a rule could be *Product = Lemon Juice* → *Store = WalMart*. The interpretation of this rule is the following: The user buys his lemon juice at the store Wal-Mart.
Although the authors approach is interesting we are not going to investigate further into it. The reason is that we look at web controlling as a control loop where goals are set and the results are checked against them. This is a holistic approach. Customer profiling is a selective view, which describes one perspective: How the user behaves.

3.5 eMarketing

Büchner et al. [BüMu98] introduce a new way to generate valuable marketing information from e-commerce shops based on web usage mining. The basis of the concept is the definition of an n-dimensional web log data cube. The cube is a repository for analytical activities. Büchner et al. concentrate in their research on pattern discovery. In order to do that, they prepare the data as following: A schema based on relational calculus is modelled in order to support analysis procedures, such as OLAP and data mining. Each dimension is represented as a relation which is connected to a fact table. Fact tables act as connecting element in a data model representing keys and summarization information. A star schema is sufficient for one given set of scenarios. For web mining, a snowflake schema is required which supports different granularities. Snowflake schemas provide a refinement of the star schema [BüMu98]. The next step after generating the snowflake schema is the analysis of the data and its interpretation. At this stage, marketing experts are involved in the interpretation of the data. These experts interpret according their expertise on the data and take appropriate actions for eMarketing events. The approach of the authors focuses merely on online shops. This is also reflected by the data sources, which are divided in marketing data, web meta data, and server data. Based on these data sets and the data preparation with hyper cubes, the results are

[7] See [EdMa97].

tailored to online shops. This is not the approach for our work and we will not focus on it further.

3.6 Benchmarks

Riedl [Ried01a] introduces a framework to measure real world workload. In his article, he looks at web servers, information retrieval systems, and transaction processing in databases. For an in-depth analysis, we refer to [Ried01a]. We will not dive into the topic of workload in our thesis because it deals with infrastructure, networks, and hardware. A working infrastructure is a prerequisite for our work.

3.7 Conclusion

This chapter has outlined the usage and application of web usage mining in different areas. From the above introduction, we see that all these research areas deal with parts of the Internet or Intranet respectively. The main difference between those and our research topic is that we do not try to improve parts of the Intranet. We aim to implement a holistic approach to the improvement of the Intranet. Furthermore, the Intranet is seen as an information space. This information space serve as an information medium for the information distribution in the company. In summary, we strive to implement a process wich facilitates the improvement of the information space.

4. Context of Research

This chapter situates the thesis in the research area of web controlling. Our approach to this topic is based on the research of R. Riedl, who introduces a framework for an adaptive, user-behaviour-reactive design. From his results, we derive an integral view to the problem of web controlling.

First, different definitions of web controlling are given. Afterwards, the chapter abstracts and discusses the various approaches to the topic among experts. Next, our definition of web controlling is given and explained. Finally, our approach of web controlling is specified and discussed.

4.1 Introduction to Web Controlling

The thesis is settled in the research area of Web Controlling. In the literature various definitions can be found for Web controlling, such as: *Web Controlling gathers the entire value chain from planning and controlling to the analysis of the whole e-commerce activities* [Syst02]. *Web Controlling is a continuing process where the actual key figures of the online activities are determined and compared to the demanded key figures. Afterwards, the differences are analysed and interpreted so as to derive appropriate actions to achieve the objectives* [Leib03]. *Web Controlling is the systematic analysis of web sites* [Hein01]. We define Web Controlling as a *continuing process of defining objectives for the usage of the information space, controlling the activities, analysing the results and taking appropriate actions to reach the first defined objectives*. In comparison to the other definitions our is close to the that of Leibrandt [Leib03]. The difference is that we include the definition of objectives in the process because we believe that this is also part of the Web Controlling. Heine's [Hein01] is too narrow and too generally defined. Systor [Syst02] looks into value chains of e-commerce shops.

Mayer et al. ([MaBe01]) point out that Web Controlling is of increasing importance but they do not give an integrative approach of how to conduct it successfully. Instead, they refer to web log mining as the controlling instrument. They state that the generated results from log mining should be processed further into an operational or strategical control system and this is a sort of controlling. They do not give concrete steps how to do that. In our opinion, the contribution is too superficial concerning Web Controlling. The main focus of the authors is web log mining. We think that this is only one part of Web Controlling and must be developed further.

Gaul et al. ([GaSc02a]) concentrate their research mainly on recommender systems. They describe recommender systems as a generic term of software, which gathers and aggregates information about web site visitors and associates it with already saved knowledge about available or obtainable offers, as well as other relevant news in order to generate recommendations for the visitor [GaSc02a]. The statement of the authors is that recommender systems are not only usable for recommendations, but can be applied for web controlling tasks as well. With their help, the main task of web controlling would be to facilitate the transformation of the visitor to a re-purchaser. Gaul et al. ([GaSc02b]) use

recommender systems resp. web controlling in e-commerce applications to increase the sales of goods of the supplier. Gaul's approach focuses on the relationship between company, marketing or sales resp. and the customer. We may refer to it as an outside view. Unlike Gaul, we focus on the internal relationships of a company. We may refer to it as an inside view.

Huber et al. ([HuSä02]) use the following approach to Web Controlling. As a basis for their concept they take the balanced scorecard as developed by R. Kaplan and D. Norton in 1996 ([KaNo96]). From the strategy and company vision, strategic goals, also called Key Performance Indicators (KPI), are derived and engaged with key figures. The KPIs represent the perspectives which are important for the economic success of the company.
Huber et al. transfer in their concept, Kaplan's perspectives into an e-business view. According to the balanced scorecard perspectives, the authors define the web scorecard perspectives as follows:

- *The system perspective*
- *The offer perspective*
- *The customer perspective*

This approach also focuses on an external view. The authors give importance to the relationship between company and customer.

Another approach to web controlling is given by Riedl [Ried01b]. Riedl proposes a framework for an adaptive, user-behaviour-reactive design of Websites based on web server log files analysis. Riedl has analyzed different web server log files and detected similar statistical patterns, which hold to Heaps' and Zipf's Law [Ried01c]. These findings lead the author to a stochastic model for Intranets and Websites and to the conclusion that the core part of the framework should be a Markov model. The Markov model represents the user behaviour on two levels: The concrete level of page impression and the abstract level of predicates. Furthermore, the author introduced a valuation function based on the Markov model with which we are able to evaluate user behaviour. This evaluation gives a feedback on how the user navigates through the website or Intranet. This feedback can be compared against the company's goal and lead to a redesign of the website or Intranet, if the goals are not achieved. Hence, Riedl's concept introduced a feedback loop for redesign of information spaces based on the Markov model.
Riedl [Ried01c] presents further a generic framework for event mining in virtual markets. The framework supports customer profiling, evaluation of websites, composition of market knowledge, measure of market dynamics, and the derivation of social role structures for information societies. The framework is based on the discovery of correlations between *a priori* and *a posteriori* observables of events within a virtual market. A virtual market is a trading platform, online shops, and other e-commerce applications. The prerequisite of the author's approach is that the virtual market is built by the component architecture principles and that each component of that market has an interface for the externalisation of services, which is an interface for issuing requests for services and is also capable to generate requests by itself. The virtual market is divided into similar entities composites into logical units, so called components. Two examples given by the author are:

1. A three layered component model (Customer Layer, Tools Layer, Data Resources Layer) can be derived from Intrawebs for document-based knowledge management or from marketing websites.
2. An e-broker architecture is divided into a four-layered component model: The customer layer, the agent layer, the request term layer and the goods layer.

Due to the component architecture, the components interact with each other by means of events. The three types of events are: Internal events, send and receive events, and remote events.
These events that are described by an event model, introduced from the author, will be discussed here shortly[8].
An event is created when one component requests a service or a resource from another component. Attributes are assigned to each event. There are two types of attributes:

1. Attributes relating to the source component issuing the request which represents the context of the request.
2. Attributes relating to the target component of the requests which represents the content of the request.

The first type of attributes is called *a priori observables*, and the second type is called *a posteriori observables*. Further definitions given by the author will omit here. Applying mathematical algorithms such as reference matrices, inference networks, Markov models, footprint growth function to the a priori and a posteriori observables, lead to customer profiles, user behaviour, product recommendations, a measure for system dynamics and an optimization of information retrieval and filtering.
The concept of the thesis takes up Riedl's feedback loop as described in [Ried01b] and develops it further. In addition, his approach of event mining ([Ried01c]) which leads to a method of analysis of the information space, is integrated into the generic filter pipeline and will be discussed in the thesis.

4.2 Conclusion

We state that web controlling is more than *"…a continuing process where the actual key figures… [are] compared to the demanded key figures"*[9]. We understand that web controlling is necessary as a user needs driven controlling process. This implies that we do not only compare figures against each other, but incorporate the user-oriented view. Latter states the paradigm shift because it gives the input requirement which must be achieved. This is expressed with the implicit annexation of the "*objective to be achieved*" of our definition. We apply web controlling to information spaces, whereas we take up the position that our approach has an integral view on controlling.
This approach is supported by the generic filter pipeline which evaluated the user behaviour and enables to compare the results against the defined objectives.

[8] For further information see Riedl [Ried01c].
[9] See [Leib03], p. 20.

5. Organizational Communication

This chapter introduces organizational communication. This introduction is necessary because the thesis deals with the Intranet as an information and communication media. The purpose of the Intranet is to instruct, to inform, and to motivate organizational members.[10] Therefore, a general overview of organizational communication is given. Different definitions of organizational communication are discussed and the difference between organizational and corporate communication is explained. Furthermore, it answers why communication is important to enterprises. Different arenas of corporate communication are introduced with their users, the stakeholders. Next, these stakeholders are added with further users of the information space. Third, the internal organizational communication is discussed. We have pointed out what the goals are and which structures and processes it has. Finally, corporate intranets are discussed and communication barriers identified.

5.1 Introduction to Organizational Communication

Organizations are social systems which are based on communication [Luhm85]. Communication is, according to Luhmann ([Luhm85]), the important element of social systems. Furthermore, they are forms of social networks which humans create to solve problems and satisfy their needs. The relation between organizations and communication processes is called organizational communication. It builds the basics for both the organization and the management and decision processes. According to Saxer ([Saxe99], p.27), the main tasks of organizational communication is to help realize the organizational purpose and at the same time, to satisfy the needs of the organizational members. Wever ([Weve96], p.167 et seq.) sees in communication the main goals of achieving agreements between two parties and in the ideal case, even similarities between them. He emphasizes that not only that the transmission of information is important, but also the understanding of the message. It is to ensure that feedback is a basic element of communication. The interviews confirm that enterprises have implemented feedback mechanism.[11]
Organizational communication is *the social glue that ties members, subunits and organizations together. Communication underlies most organizational processes, contributes to both the development and the individual characteristics, including size, department, autonomy, and upward aspirations. Without communication, organizing could not occur* ([EuRo87], p. 42). The results of communication are the goals which the organization pursues ([Mast02]). In the literature, different definitions are given for organizational communication. For Bruhn ([Bruh97], p.1), communication is *the transmission of information and content to manipulate opinions, attitudes, expectations and behavior patterns of the addressees according to a certain purpose*. Theis-Berglmair ([Thei03]) defines organizational communication as *communication in organizations and from organizations*. Szyszka ([Szys99]), on the other hand, defines organizational communication slightly broader. In his view, it is *communication in, from and about organizations*. Stohl ([Stoh95]),

[10] See chapter 6.2, point 10, p. 58.
[11] See chapter 6.2, point 15, p. 62.

to give another, defines it as *the collective and interactive process of generating and interpreting messages* ([Stoh95], p. 5). Grunig ([Grun92]) explains organizational communication from a public relation point of view, stating that organizational communication is public relation in a narrower sense. Public relations are the *management of communication between an organization and its publics*. Public relations describe *the planning, execution and evaluation of an organization's communication with both external and internal publics*. Publics, therefore, represent the *groups that affect the ability of an organization to meet its goals* ([Grun92], p. 4). Although there are many definitions for organizational communication, they all have the same topic – the communication relation between the internal and external environment ([Mast02], p. 11). As Grunig ([Grun92]) stated, organizational communication deals with external and internal stakeholders[12]. In chapter 5.3 we give a more detail overview of company's stakeholders and the interaction among each other.

Categorically, external and internal organizational communication can be differentiated. The external communication deals with the social and political environment of the company and the economical and technical environment ([Mast02], p. 13). On the one hand, it initiates and manages the process of coordination and agreements of interest in the market place which are economical relations with suppliers, customers or competitors. On the other hand, the external communication deals with the sociopolitical environment which contains the uneconomical activities and publics such as politics, science and art ([Zerf96a], p. 289 et seqq., [Zerf96b], p. 38 et seqq.).

Since we focus on corporate Intranets in our thesis, the main topic we are interested in is the internal communication of organizations. This means that we will only look as far as it is necessary for our understanding into external organizational communication. As we have mentioned above, enterprises are our organization[13] of interest which leads us to the definition of Mast ([Mast02]). She defines corporate communication as organizational communication where the organization is a company ([Mast02], p. 7). Corporate communication comprehends all communicative processes between the company and its members ([Mast02], p. 241). Zerfass defines corporate communication as *being all communicative activities of organizational members which contribute to the labor definition and completion of a profit-oriented business entity* ([Zerf96a], p. 287).

Internal communication is on the one hand, a constituent communication which a company needs to get set up, to obtain a constitution, to attain an identity and become a definable economic actor. On the other hand, internal communication assures the goods and services through coordination and division of labour of the organization insiders. Furthermore, it consists of the communication behavior of the organization members to achieve the company's goals ([Zerf96b], p. 38 et seq.). *Organizing communication is tantamount to organizational leadership. It is part of the strategic management of a company because communication connects on the one hand the management level with the operations level and on the other hand the different communication directions on the same level* ([Szys99], p.16 and p. 27), which enables the company to act profit-oriented in its economic environment.

As Dyllick et al. [DyMe02] state, wherever persons interact together, communication takes place. Communication is not limited only to verbal exchange between two individuals. It also

[12] The word "stakeholders" has replaced the word "publics".
[13] Unions, parties, etc. are also organizations ([Szys99], p.6).

takes place when it is denied[14] and furthermore involves also gestures, mimic and manifold other actions which are perceived and interpreted by others. For Luhmann ([Luhm85]), arguing from a system-theoretical point of view, communication consists of three elements: Information, message and understanding. It is accomplished in that degree in which the information sent is understood. A forth element exists that the receiver of the information, after having understood it, either accepts or rejects it. That is, he can either act according the senders' expectations or against. Based on the above mentioned remarks, Dyllick et al. [DyMe02] define communication as follows:
Communication is considered as a goal-oriented, verbal and non verbal elements comprehend exchange of information between individuals which aims to manipulate the receivers.
The authors state that strictly speaking, there is a dependency relation between two different communication goals, manipulation, and understanding. The primary goal of manipulation can only be achieved if the secondary goal of understanding succeeds.

McQuail et al. [McWi93] cite Lasswell's famous phrase of communication process (s. figure 9).

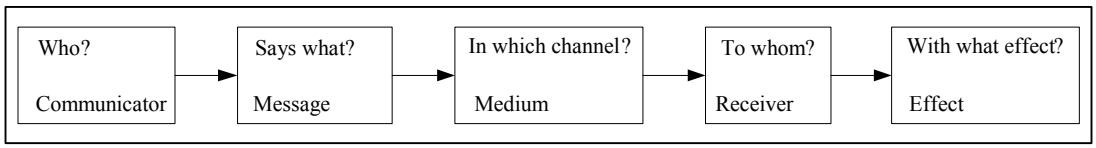


Figure 9 Source [McWi93], Lasswell's formula of communication process

They also present in contrast Braddock's extension of Lasswell's formula (s. figure 10). In his model, Braddock adds two more aspects of the communication act. The circumstances under which the message is sent and for what purpose it is sent.

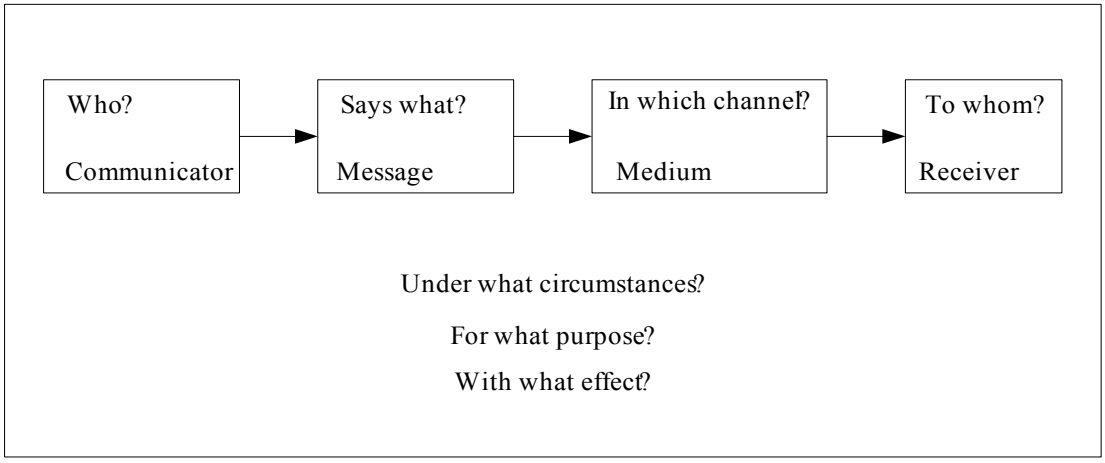


Figure 10 Source [McWi93], Braddock's extension of Lasswell's formula

A communication process, therefore, consists of at least two actors, certain information content, and a certain communication media as well as a particular effect [DyMe02]. Thereby, the immediate communication (face-to-face) in the form of speech and characters is called *direct communication*; the one conducted through technical media (e.g. Intranet, Extranet, print media) is called *indirect communication*.

[14] "One can not not communicate", [WaBe03], p.53.

Dyllick et al. distinguish between three communication models. In the context of a technical communication model, the medium with its possibilities and limits of information transmission plays the central role. Each medium has its own performance characteristic. Direct communication over the medium "speech" reaches only a few receivers, but is able to transmit rich content and allows true communication (*two-way communication*). Indirect communication is able to reach a larger amount of receivers, but with the loss of richness of content and mostly limited to pure information (*one-way communication*). Evans et al. [EvWu00] define reach and richness as follows: Reach is the number of people exchanging information. Richness applies to six aspects of information:

- Bandwidth or the amount of information that can be moved from sender to receiver in a given time
- The degree to which the information can be customized
- Interactivity: In a small group a dialogue can take place, but to reach a very large group the message can only be a monologue
- Reliability: Information is reliable when exchanged among a small group of trusted individuals but is not when it is circulating among a large group of strangers
- Security: Highly sensitive is only shared with a small group of trusted people, less sensitive information is also spread to the public
- Currency: Information can be received instantly or delayed

In their contributions, they determine the relationship between richness and reach. They state that the more reach the information has, the smaller the richness of information is and vice versa. This holds especially for traditional media. With the emerging of the new media Intranet, the traditional compromise between richness and reach is broken open. It creates the preconditions for a comprehensive, individualized communication and interaction with a large number of widely distributed persons ([EvWu00]). From the outcomes of the interviews, we see that the content of the Intranet contains department specific information, news, corporate and product information, message board, etc. This is an example of the breakup between rich and richness.[15]

In the context of a psychological communication model, the main focus lies in the perceptions, interpretations, opinions and the values of the communication partners. These are rather implicit than explicit and rather subjective than objective. Whether communication is successful or not, is decided at the receiver's site. These successes are determined on the basis of how messages are perceived and interpreted, what relevance is attached to them and what they provoke depends on their pre-comprehension and expectations. According to Watzlawick et al. ([WaBe03]), each communication has a content and a relation aspect. The content aspect transmits the information, and the relation aspect determines how it should perceived.

The sociological communication model is focused at the collective communication structures. Individual perception and interpretation structures are no longer seen as independent, but rather as dependent on comprehensive, social constructed communication structures, which have emerged and been consolidated as valid perceptions in the society or in sub-areas thereof.

[15] See chapter 6.2, point 3, p. 51.

5.2 Arenas and functionality of corporate communication

Corporate communication supports the exchange with internal and external stakeholders. Depending on the stakeholder, the communication takes place in different arenas, with different actors, contents, goals and conditions. Dyllick et al. [DyMe02] differentiate four communications arenas as follows (s. figure 11): (1) employee communication, (2) market communication, (3) investor relations and (4) public relations.

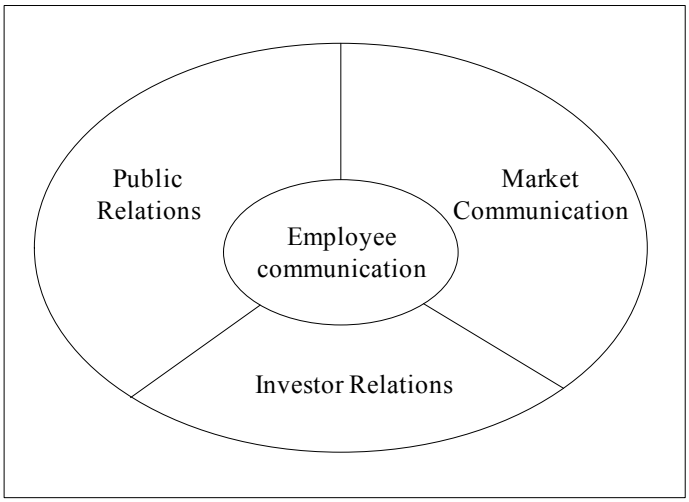


Figure 11 Source [DyMe02], Arenas of corporate communication

We will give a short overview of the communication arenas as described in Dyllick et al. [DyMe02].

Employee communication

Employee communication includes all communication between the corporate insiders. It contains the coordination of strategy and organization[16], the emerging of a common corporate culture, the reengineering of business processes, the restructuring of business structures, as well as wage bargaining. Employee communication has an understanding and coordination function in the company. It supports the social integration, the structuring of business processes and tasks, as well as the management and the coordination of actions ([Zerf96]).

There are formal and informal types of communication in the company, hierarchical communication between supervisor and employee and lateral communication between co-workers. Furthermore, there is direct (discussions, workshops, etc.) and indirect communication (letters, documents, etc.) between the company members. Employee communication mainly has an *understanding and coordination function* and is essential for the goods and services output of the company. It is a management tool and the feedback

[16] See [KeRi05], strategy and organization support are named by the interviewees, see point 3 and 10, p. 51 and p. 58, resp.

between the communication partners is of vital importance. The interviews show the most common feedback mechanism such as email, contact form, etc.[17]

Market communication

Market communication considers customers, suppliers, and partners as external stakeholders of a company. Therefore, it is focused on the buying and selling market and gives special attention to the goods and services of the company and the consumer needs. The goal of market communication is to manipulate knowledge, attitude, and behavior of existing and potential customers so that latter will favor the company's products over those of the competitor's. Market communication has a *persuasion and manipulation function*. Its task is to support the business process in the area of market and service development.
The most important tools of market communication are advertising and sales promotion, mostly conducted over indirect mass communication media. It is either over electronic (TV, Internet, radio) or print (newspapers, posters) media.

Investor relations

The function of investor relations is *trust creation and expectation management* for the capital markets. Its task is a regular preparation and annotation of financial and strategic information and also forecasts of turnover, market shares, cash flows for the major investors, financial analysts and journalists.

Public relations

Public relations deal with the role and function of the company inside the society. Company activities often affect the public interests such as layoffs, ecological and sanitary topic. Thus, the company must deal with various public stakeholders, such as government, politicians and residents. The function of public relations can be seen in the creation of *understanding and legitimation* in cooperation with relevant stakeholders.

5.3 Stakeholders

In this section, we are going to determine the stakeholders of the information space and discusses the interaction with one another. Figure 12 gives us an overview.

We identify eight stakeholders: Management, Investor, Customer, Employee, Supplier, Public, Government and Strategic Partner. We are not going to consider all stakeholders because some will not really communicate through the information space. The first stakeholder who will not be considered is the public because the communication to it is limited to the annual report, brochures, websites and media on the company's activities. The second stakeholder we will not consider is the government because there is no interaction between it and the information space.

[17] See chapter 6.2, point 15, p. 62.

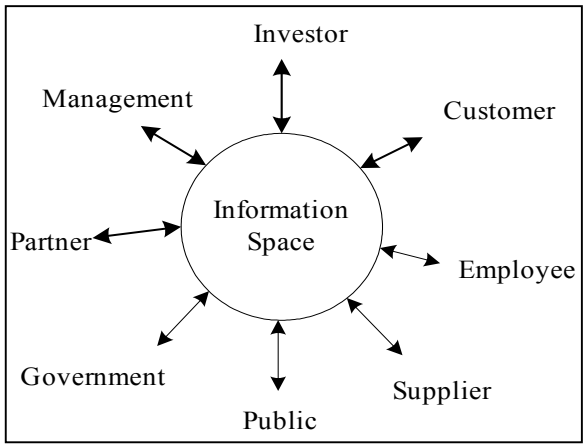


Figure 12 Stakeholders

In the following section, we are going to describe the interactions between the other stakeholders and their usage of the information space. Before we start, we first look closer at the stakeholders. There are two kinds of stakeholders: The ones who are employed by the company (Management, Employee) and the ones who are not employed, but are in some way related to the company (Customer, Supplier, Partner, Investor).
The information space can therefore be divided into two parts: The Intranet and the Extranet. The Intranet is used by all employees of the company, i.e. the corporate insiders. The Extranet is used by the stakeholders who are related to but not employed by the company, i.e. the corporate outsiders.

Management

First, the management uses the information space to communicate with the employees. The main usage is to put directives and instructions on and to conduct survey through it. Second, it is also used to communicate within the management team. Third, the information space is also a communication channel to the company's partners.

Employee

The employees use the information space as a communication channel to put information of ongoing and past projects on it, to exchange knowledge, or as an archive where any kind of information is stored. Furthermore, employees also talk to suppliers and customers. This can be a request for installation guidance from a customer or a technical support of an employee to a supplier.

Customer

Customers access the information space through the Extranet. They use it to download their reports and agreements, to purchase products or to communicate, according to the company policies, and to the employees for e.g. technical guidance.

Supplier

Suppliers also access the information space through the Extranet. They upload their new price lists, new warranties and agreements, and are able to see when their costumers need to be provided with new supplies.

Partner
The management also needs to communicate with their partners. This is the case when companies have joint projects. They exchange documents, contracts and project status. The partners also have the possibility to communicate with each other through the information space of the company. In our interview, an automotive manufacturer[18] stated that their commercial partners, i.e. the car dealers, have the possibility to exchange information with each other and therefore improve the product and service quality for the customers, as well as point out the weaknesses of the product. From these experiences, the automotive manufacturer is also able to benefit in that way, as the information flows back into the product development and quality assurance.

Investor
Investors get financial documents from the company. In most cases, the company has implemented an Extranet to allow the investor to access the necessary documents and eventually download them.

The columns and the rows of table 1 show the stakeholders. The cells from the table indicate through what channel of the information space (Intranet or Extranet) the stakeholders communicate with each other. No entry means that there is no communication between those two stakeholders.

	Management	Employee	Customer	Supplier	Partner	Investor
Management	Intranet	Intranet	-	-	Extranet	Extranet
Employee	Intranet	Intranet	Extranet	Extranet	Extranet	-
Customer	-	Extranet	-	-	-	-
Supplier	-	Extranet	-	-	-	-
Partner	Extranet	Extranet	-	-	Extranet	-
Investor	Extranet	-	-	-	-	-

Table 1 Interactions through the information space

5.4 Internal organizational communication

Since the organization acts accordingly to its organizational purpose, it is forced to manipulate its members in a certain way in order to achieve its goal. The organizational purpose on the one hand is the production of product or services, and on the other hand, it is the guarantee of continuity of the organization ([Szys99], p. 7). Along the lines with Hoffmann ([Hoff01]) and according to the definition of Zerfass ([Zerf96a]), the goal of internal communication is the achievement of the company's objective. This is ensured by providing relevant information to the employees as shown by the interviews. Figure 18 list "corporate goals" as a topic of the Intranets content. Excluded from internal communication are those communicative activities which do not support those objectives. Rommert ([Romm02], p.101) states that the aim of internal communication is to improve the company's performance. Armbrecht ([Armb92], p. 296) also points out the increase of the

[18] See [KeRi05], Interview with Automotive 1, p.70.

company's performance which is aligned with the company's goals. This statement is also confirmed by the interviews. Figure 24 lists "efficiency, effectivity and quality" as a benefit which is achieved by the Intranet usage. Gryza et al. ([GrMi00], p.12) see in the increased international competition and the demanding requirements of the company that they have permanently adapt to the changing environment internal communication, thus becoming an important competitive factor. For Mast ([Mast02], p. 244), communication is important for the company if it wants to secure competitive advantages in the market place because communication enables fast acting. Here as well, we can confirm from the interviews that enterprises use the Intranet for fast information distribution and information access to increase competitiveness.[19]

Internal communication serves to motivate the employees, increases the employee performance and fosters the corporate identity. This statement can also be confirmed as well by the interviews. Figure 23 and 25 list the "efficiency and effectivity improment" and the support of "corporate culture". Furthermore it supports the formation of employee opinions in the company and helps to understand necessary changes of the company ([Armb92], p. 4 and p. 302 and [Mast02], p. 244). Point 10 of the interviews confirms this statement. The employees are integrated in the company's decisions and therefore feel as a part of the company. Mast ([Mast02], p. 244) also emphasizes that internal corporate communication improves the implementation of corporate goals at all levels of the company and optimizes the communication architecture.

5.5 Employee Communication

In the section above, we have given an introduction on internal communication. In this section, we will give a more detailed view on this topic.

Hoffmann [Hoff01] classifies communication in companies as task-oriented and social functions. The latter is concerned with the relation between the company insiders, and the former is focused on task performance and goal implementation. Main functions of communication are the coordination of action and the supply of information which are needed in order to accomplish the required tasks.

The tasks of an organization which must be performed are usually too large to be handled by a single individual. They have to be split up and it must be decided which sub-tasks must be executed by each organization member. The specialization creates complexity in the organization and with the distribution of the task to a number of employees, the individuals are no longer able to overlook all activities. Moreover, interactions are created between the organization members due to the division of work. At the end, the solutions of the sub-tasks must be merged together and be coordinated. Problems of the specialization can be reduced through the use of instances with decision and management functions and also with the use of communication instruments for coordination. The effective realization of specialization and coordination needs communication structures and the exchange of information between the organization insiders.

Control instruments may be directives, self-determinations between the co-worker, programs and plans. Directives are hierarchy-bounded, and therefore a vertical communication

[19] See chapter 6.2, point 7, figure 20, p. 55 and point 12, figure 24, p. 59.

direction is appropriate. Self-determination implies a communication between coequal co-workers. This communication process demands a horizontal communication direction.

A further function of the employee communication is to create a social relation in the organization. The employees are integrated in the organization through the values, norms and corporate culture. The usage of the Intranet as information channel should also support the social life in enterprises. This statement is made by the interview partners. They mentioned that the Intranet should support and foster corporate culture, and be a forum for personal interaction in the company.[20] Communication in companies satisfies the social function of creating and sustaining a social structure. The social factors of corporate communication are very important for the functioning and the success of a company because they create social relationships which allow the cooperation. In addition, the social relationship can act as a motivation source. At the least, the success of leadership depends on the social relation between supervisor and employee.

5.5.1 Information goals of the company and the employees

The tenor of the management is of significant importance for the employee communication goals. From our interviews, we have derived following communication goals.[21]

- Comprehensive employee information
- Fast communication
- Intranet becomes a work instrument
- Up-to-dateness of information
- Communication channel of the management
- Efficiency & effectiveness improvement
- Transparent communication
- Support organization
- Support business
- Open communication

Macharzina ([Mach90], p.63) distinguishes between following characteristics of corporate goals:

- Employee orientation: There is a right to information for the employees which is of legal nature and derived from the close relationship between the employees and the company.
- Company orientated: Information activities are seen as instruments to realize economical goals.
- Employee and company orientation: The interests of the company as well as those of the employees should be considered and the transparency of the company should be fostered.

[20] See chapter 6.2, point 7, p. 55 and point 10, p. 58.
[21] See chapter 6.2, point 10, p. 58.

In his research, Macharzina cites additional empirical results ([Mach90], p.67 et seq.) where further goals of employee communications are listed. In the following. we list an extract. These goals agree by the majority with our findings:

- Integration of employee in the company activities
- Reveille employee's sense of responsibility
- Motivation of the employee
- Better participation of the employee in corporate activities
- Reduction of uncertainty of the employee
- Reduction of dissatisfaction of the employee
- Build trust
- Satisfaction of employee's information needs

Macharzina describes further the information goals of employee communication from a management point of view. In his research, he derives the categories company oriented and employee oriented goals.

Manipulation of behavior	Employee oriented
Associated characteristics:	Associated characteristics:
a. Integration of the employees Identification with corporate goals Target appropriate behavior	a. Relationships Perpetuation of personal relationships Improvement of the relationships within the company
b. Job fulfillment in the company Motivation Quality assurance Better job performance Improvement of cooperation Creation of a sense of responsibility Creation of quality awareness	Good working atmosphere Publication of operational activities Creation of a sense of togetherness Creation of trust Understanding of other work areas Empowerment of the employees
c. Formation of opinion Transparency of operational decisions Comprehension of operational decisions Fostering of entrepreneur thinking	b. Personality of the employee Show the importance of the employee Self-conception for the work Building a feeling of security Decrease the isolation of employees Employee satisfaction
d. Representation of the company Inwards public relations Outwards representation of the company Image creation/ advertising effect	Satisfaction of information needs of the employees
e. Conflict management	c. Employee development Improvement of employee qualification Development of sociopolitical critical

Conflict avoidance Conflict diminution	faculty Creation of a competitive information edge of the employees

Table 2 Source [Mach90], Dimension of the information goals of the management

Conducted from a survey, following employee communication tasks were derived by Sorg [Sorg79].

- Fulfillment of corporate tasks
- Identification with the tasks and the company
- Motivation to perform
- Promotion of one's initiative and the sense of responsibility
- Transparency of the corporate activities
- Integration into the company
- Information as leadership instrument
- Promotion of the sense of togetherness
- Improvement of the internal teamwork
- Understanding for the leadership
- Improvement of the working morale
- Improvement of the working atmosphere
- Decrease of fluctuation
- Increase of individual state of knowledge
- Satisfaction of personal motives and needs

The information goals must be institutionalized through the information principles. Those are according to Macharzina ([Mach90], p. 77):

- Clearness
- Truth
- Up-to-dateness
- Receiver orientation (formal and textual)
- Understandability
- Regularity
- Objectivity
- Free of contradiction
- Information reach
- Information richness

Picot et al. [PiRe87, p.47] state four required principles for each organizational communication:

- Rapidness/Convenience: Short transmission time, short preparation time, fast feedback, convenience of the transmission process.
- Complexity: Problem solving, clarification of controversy.
- Confidentiality: Transmission of confidential content, authentication of the receiver, protection of message corruption.
- Correctness: Transmission of proper wording, high volume of information, simple further processing.

Tonnemacher [Tonn98] further mentions that an informed employee is motivated and performance-oriented. He feels as an integral part of the company and identifies with it.[22] The information goals by Tonnemacher are the following:

- Information need: Orientation, learning, to satisfy curiosity
- Need for personal identity: Self-discovery, reinforcement of personal values
- Need for integration and personal interaction: Social sympathies, affiliation through identification, basis for conversation and interaction
- Need for entertainment: In the company of limited value, emotional discharge, escapism, edification.

Burkhart et al. [BuHö92] present four communication models for public relations. These can also be applied to internal communication. The models differ in the communication direction (one-way vs. two-way communication) and in the communication comprehension (symmetric vs. asymmetric communication). The *Publicity* model describes a propagandistic company which appeals for a positive profile in the public. The *Information activity* model distributes correct information to the publics without being concerned about the effect resp. feedback. The *Asymmetric* model describes the intention to influence target groups and to gather systematically the reactions. Finally, the goal of the communication of the *Symmetric* model is to attain a better understanding for the problems of the involved persons.

Characteristics	Models			
	Propaganda	Information activity	Asymmetric communication	Symmetric communication
Purpose	propaganda	distribution of information	persuasion by means of scientific insights	mutual understanding
Type of communication	one-way; complete truth not relevant	one-way; truth is important	two-way; unbalanced impact	two-way; balanced impact
Communication model	sender → receiver	sender → receiver	sender → ← receiver feedback	group → ← group

Table 3 Source [BuHö92], Four models of public relations

5.5.2 Communication structure and communication process

Communication is understood as a process at which information is exchanged for the purpose of a task oriented understanding. It has two functions: A textual (transmission of factual information) and a social function (development of a personal relationship) ([PiRe87]). Communication processes can be arranged through various criteria. Picot et al. ([PiRe87]) identify characteristics according to:

[22] This statement is approved by the interviews, see [KeRi05].

Characteristics	Value
The receiver of an information	Individual – mass
The direction of communication flow	One-way – two-way
The dialogue mode	synchronous – asynchronous
The organizational level	Vertical – horizontal
The structure of the transmission process	Single-level – multi-level

Table 4 Extract from source [PiRe87, p.38], Characteristics of communication processes

Bruhn ([Bruh95]) lists following main criteria:

- type (personal, impersonal communication)
- intensity (intensive, insistent, passive communication)
- frequency (one-time, repeated, sporadic communication)
- impact (cognitive, affective, conative reaction of the addressee)
- degree of cooperation (one-way, two-way communication)

He distinguishes the relevant communication processes for corporate communication between the following two criteria: (1) the medium of communication and (2) the direction of communication. The medium of corporate communication can be the management or the employee. The direction of communication is further divided into three different forms: (1) downwards directed communication (2) upwards directed communication (3) sidewards directed communication. Table 5 shows the types of communication processes.

Addressee / Sender	Management	Employee
Management	Corporate interaction	Employee communication
Employee	Corporate oriented employee dialogue	Employee interaction

Table 5 Adjusted from source [Bruh95], Types of communication processes

The downwards directed communication is communication from management to the employee, where the management influences employee's behaviour through orders. The upwards directed communication on the contrary is from the employee to the management. This communication is a form of reporting. The sidewards directed communication is communication between individuals of the same hierarchy level. Anders ([Ande83]) as well as Bartram ([Bart69]) have also introduced three communication process directions, where they name them vertical, horizontal and diagonal communication. The vertical communication runs in two directions; from the management to the employee and vice versa. Former is also known as the instruction path, latter as the reporting path. The horizontal communication is an information exchange between members of the same hierarchical level. Diagonal communication takes place between members of different hierarchical levels and different organizational units. Greenbaum et al. ([GrHe95]) introduced the same three communication processes where they name them: Upwards, downwards and horizontal. Anders divides communication processes further according to their communication level. Processes where the sender and the receiver are on different hierarchical levels can be single-level or multi-level resp. direct or indirect. If there are no further hierarchy levels between

the sender and the addressee, or if they can be jumped over, then a single-level communication process exists. In contrast to that, there are several hierarchical levels between sender and addressee at the multi-level communication process.

As a further characteristic of organizational communication, Anders ([Ande83]) mentions the communication unit. He distinguishes between four communication units: The intrapersonal communication, the interpersonal communication, the organizational communication, and the mass communication. The interpersonal communication deals with communication processes within an individual and describes the reactions of an individual at information reception. The interpersonal communication is divided into three sub-categories: The dyadic, the intra-group and inter-group communication. The first is the communication between two persons. Second, it is the communication within a group which is called intra-group communication. Groups can be permanent established such as business units or departments, or they can be temporary established, such as project teams. Finally, communication between different groups is called inter-group communication. The third communication level is divided into the intra-organization and inter-organization communication. The dyadic, intra- group and inter-group communication form the intra-organization communication. Therefore, it describes the communication within an organization. In contrast, the inter-organization communication deals with communication between different organizations. Mass communication is understood as communication between a single person and an undefined number of persons. Another characteristic of organizational communication is its communication purpose. It describes the effect which should be realized at the addressee, such as instruction, information, or motivation.

Anders ([Ande83]) also names the communication roles organization members can obtain. These are: *Star, Liaison, Isolate, Bridge, Initiator, Relayer, Boundary spanner and Terminator*[23]. Stars can be at the same time Liaisons, Bridges, Initiators, Relayers, Terminators and Boundary spanners. These role takers take active part in the communication process of the organization, where the Isolates are passive and with no relationships to other organizational members. Social networks are another part of communication structures within the company. Monge ([Mong87]) distinguishes between personal networks, group networks, organizational networks and interorganizational networks.[24] Personal networks are communication linkages which people maintain with other individuals. Group networks describe the patterns and structure of people who communicate more with each other than the rest of the people in larger networks. Organizational networks represent the structure differentiation of organizations. In such networks, different network roles can be identified such as Star or Liaison.[25] Interorganizational networks are the configurations of communication relations between organizations.[26] In his experiments, Bartram ([Bart69], p.124 et seqq.) asked five people to solve a problem together. He notices the emergence of three communication networks: A complete network, a circle and a star. The complete network allows that everyone of the network is able to communicate with each other. In the circle, the communication of member is limited to his immediate neighbours. The star allows one member to communicate with the other members, but latter have no communication

[23] For a more detail description of the roles see [Ande83], p. 40.
[24] [Monge87], p. 241ff.
[25] See also above regarding Anders.
[26] See [Mong87], p.242.

possibility to communicate with each other. This implies that the person in the centre of the star has a powerful position.

Another part of communication structure is the formality of the communication process ([Ande87] and [Bart69]). Communication can be formal or informal. Formal communication results from the division of labour and is considered a task-oriented communication. Informal communication has an important purpose in the organization, too. It stabilizes the formal organization by compensating the weaknesses and making them more flexible. Furthermore, it meets social needs such as familiarity and affiliation.[27]

5.5.3 Design of organizational communication

Hoffmann introduces three components for the design of organizational communication. The components are: The communicative competences, the communication structure and the communication technology (s. figure 13). The goal of the design of organizational communication is to establish efficient communication processes with regard to functional and social relationships, communicative competences, structures and technologies ([Hoff01], p. 68 et seqq.).

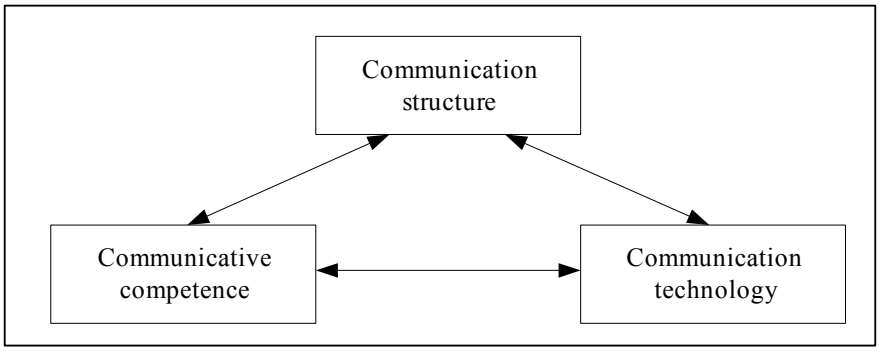


Figure 13 Source [Hoff01], Design of organizational communication

Communicative competences

In order to communicate, individuals need communicative competence. Communicative competence is understood as the general faculty of speech of individuals[28] or the capability to show behavior which is not imitated[29]. The ability to communicate with each other is a prerequisite for employees in a division of labor based task for the achievement of objectives, and is also important for a good communication climate. Communication in organizations is not only made through language, but also through technical media. This implies that communicative competence of verbal and non-verbal capabilities has to be supplemented with media competence. Latter is the capability of interaction with media and the understanding, perception, reception, processing und design of media content.[30] Therefore, companies have to be aware that they must provide training offers for the employees to

[27] See [Hoff01], p. 60, [Ande83], p. 39 et seq. and [Bart69], p. 136 et seqq.
[28] See [Hoff01], p. 70.
[29] See [Baac73], p. 260.
[30] See [Hoff01], p. 71.

eliminate deficits in communicative and media competence.[31] The interviews showed another picture of the reality. Hardly any company conducts employee training concerning the media competence. The argument is that its usage should be self-explanatory. Likewise, there is no training on the improvement of communicative competence. If help is needed from an employee to compose an article, then corporate communication assists.[32]

Communication structure

Employee communication takes place in the context of structural conditions.[33] The goal of communication structures is to build fast, secure, exact and comprehensive communication processes.[34] They must enable an effective and efficient coverage of employee's communication needs. In previous chapters, we have given a detailed description of the structures and processes. The company has to make the decision on what kind of structure and process fits its needs the best.

Communication technology

Information and communication technologies consist of systems for the creation, processing, storage and transmission of information. The information and communication media used in companies are computer networks, databases, personal computers, printers, business TV, video conferencing, email, etc. The potential information and communication technologies strongly influence the organizational communication. Meier ([Meie02], p.103 et seqq.) states that the introduction of new communication technologies changes the work processes, lead to new interaction situations and new communication forms. Our interviews not only confirmed the change of work processes, it also showed that the information and communication processes as well as the organization processes have changed.[35] The company Research Laboratory emphasizes the new communication forms such as virtual communities, conference calls and the tracking of on-line presentations. The more the budget is cut, the more new communication technologies are used.[36] He describes video conferencing as a new form of communication between spatial distributed teams. Further, Krzeminski et al. ([KrZe98]) as well as Thimm ([Thim02]) show the impact of information and communication technologies on corporate communications. Within the organization, it has to be defined with which technology the employee's and company's communication needs can be best satisfied. In our thesis, we concentrate on the Intranet as an information and communication channel in the enterprises.

5.6 Corporate communications by means of the Intranet

There is not much literature about the Intranet as an information and communication media in organizational communication. Only a few works have looked into this subject. Hoffmann ([Hoff01]) analyzed the Intranet as a medium for employee communications. In the first part

[31] See [Tonn98], p. 101.
[32] See chapter 6.2, point 13, p.61
[33] See chapter 5.5.2, p. 36.
[34] See [Hoff01], p. 73.
[35] See chapter 6.2, point 2, figure 17, p. 50.
[36] See [KeRi05], interview with Research Laboratory, p. 84.

of his work, he gives a theoretical overview of corporate communications. It the second part, he represents the results of a survey that he conducted in companies in order to gather data on the use of the Intranet for employee communications in organizations. With that, he conducted interviews and written surveys. One has to emphasize that he concludes that Intranet communication changes the communication, the work situation, the cooperation and the organizational structure of a company. Rommert ([Romm02]) also gives a theoretical overview of corporate communications. His intention is to derive possible tasks from a theoretical point of view for the Intranet in a company. He identifies five tasks for the Intranet. The first task is to foster the formation of common constructions of the realness. The second task is to support the design of a system-environment-relationship. The third task is to accompany organizational development, to foster structure development, and to support process modification. The fourth task is the acceleration of processes and the last task is to foster learning processes in the company. Stürzebecher et al. ([StSc99]) consider as example the Bertelsmann AG that introduced the Intranet to foster the communication. Bertelsmann AG developed the Junior-Net, the Intranet for management trainees. The goal of the Junior-Net was the continuous support, the development of the management trainees by means of the general management development and the creation of a world-wide information network. The Junior-Net was expected to support the world-wide communication, the information exchange and knowledge transfer of the executives. Nusch et al. ([NuGu98]) cite Hoechst AG as an example of internal communication by means of the Intranet. The Intranet allows fast communication and connects world-wide all corporate groups. The information disadvantage of remote groups vanishes. Shell can be mentioned as further example ([KüMü98]). This company also used the Intranet as a communication media. One must emphasize that Shell is aware of the fact that a successful use of the Intranet as a communication media entails communicative competence of the organizational members.

5.7 Corporate Intranets

In our thesis, we treat the Intranet as a communication media in a company. Therefore, we first examined the roles and tasks of corporate communication in a firm before we derive the requirements which the Intranet must fulfill in order to satisfy them.[37] Further, corporate communication is divided into internal and external communication. Internal communication is within the management body, within the employees and between management and employees.[38] External communication deals with the relationship to the stakeholders, who are among others the customers, investors, suppliers, and the public.[39] We focus merely on internal communication, but will also touch the external.

Communication as a management instrument is very important to the company. This is also mentioned by the interview partner. They name the Intranet a communication channel of the management.[40] These days with the restructuring of companies, the layoffs, and the rapid change of market conditions, it is essential for the company's success to include the

[37] See chapter 5.3, p. 29 et seq. and chapter 5.5.1, p. 33 et seqq.
[38] See also chapter 5.4, p. 31 et seq.
[39] See also chapter 5.2, p. 28 et seq. and chapter 5.3, p. 29 et seq.
[40] See also chapter 6.2, point 10, p. 58.

employees into the company strategy and –goals ([Schi02]). The role which falls to communication is among others to facilitate the company's change processes. It should initiate the employee into the new challenges which the company faces and win him to take part in this process.
On the other hand, communication between the employees is equally important. It fosters the information exchange in a company if established as corporate culture. Thus, it motivates the employees to share their knowledge among each other and to take actively part in the company's prosperity.
From the above mentioned statements, we are able to derive some tasks for corporate communication ([Schi02]):

- Determination of communication objectives derived from corporate strategy
- Implementation and allocation of the overall communication media and – platform
- Advising and supporting the management in their communication tasks
- Controlling of corporate communication

We talked about external communication mentioning customers, strategic partners and of likewise. However, this classification is too coarse in order to separate the external employees from the internal. The work — and organizational structures are changing and the borders are becoming increasingly permeable. Let us think about the different type of workers: Temporary employees, employees from partners and freelancers. Hence, the corporate communication must also take those "external" employees into account.

We have given an overview of the tasks corporate communication has. Among others it is the *implementation and allocation of the overall communication media and – platform.*
If we look to the Intranet as a corporate communication channel, we have to ask what the tasks of such a channel are. Furthermore, whom it will serve and how can the benefit be measured?
Gryza et al. ([GrMi00]) present a survey from Siemens Information and Communication Networks. Following results were found: The Intranet's Return On Investment was 215%. Converted into the daily information retrieval effort, it was an efficiency improvement of 17%. The time saving for most employees was 18 minutes. This is a positive example which shows that the Intranet is becoming increasingly a competitive factor. Unfortunately, there are also negative examples from companies which they could not benefit from it.[41] In contrast, the Intranet grows uncontrolled, and without any concept, it results in the refusal of its usage.
New media have contributed to the change of work culture. As we introduced the different types of workers[42], there are also different types of communication which the company has to take care.

[41] As described in chapter 1.1, p. 1, Problem and Motivation.
[42] Temporary employees, freelancers and employees from partners working on site.

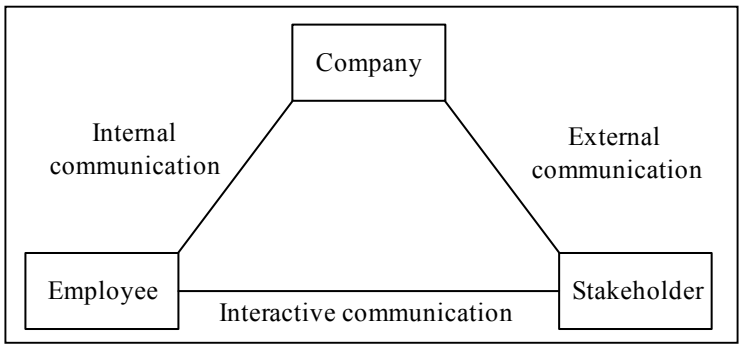


Figure 14 Modified communication model ([GrMi00])

Internal communication is between the company and the employee. E.g. the company presents to the employees the corporate strategy for the next five years and the impact which it has.
External communication occurs between the company and the stakeholder. E.g. the company announces the financial report to the investors.
Interactive communication is between the employee and the stakeholder. E.g. the company's software engineer gives direct support to the customer concerning the upgrade of the railway control system.

The above mentioned communication models are also given in the chapter "Stakeholders" and will be discussed.[43] We intended to give here a further view of different communication types.
External communication, in the context of our work, is made through Internet or Extranet. In our thesis, we are looking at the information space as a communication channel with the result that we look into the above mentioned topics.

What are the functions and tasks of an Intranet within the boundaries of communication to make it an integral part of the company, so that the users can benefit from its usage? A listing is given below:

- Easy access to information
- Standardized information culture in the company
- High quality of information
- Support the internal and external communication
- Enable information filtering

5.8 Communication Barriers

Hoffmann ([Hoff01]) cites that the barriers for an effective use of the communication channel were "Hard-Wiring"-blockades and "Soft-Wiring"-blockades. "Hard-Wiring"-blockades are the incompatibilities of the technical systems, limitation on access and error rates of the system. The "Soft-Wiring"-blockades, on the other hand, are the techno illiteracy of the users based on missing education and media competence. In particular, as one of our interview partners stated, there are deficiencies on the management level with the interaction and the evaluation of new communication media.[44]

[43] See also chapter 5.3, p. 29 et seq.
[44] See [KeRi05], Interview with Telecommunication 3, p. 33

The introduction of a new communication media leads to an increase of content quantity, but not necessarily to an increase in quality.[45] More data are distributed, which leads to an information overload of the users. Macharzina ([Mach90]) emphasizes that further barriers are the non attainment of large parts of information to the users, and some information is simply not known by the users, or the offered information is not in line with their needs. Further barriers are the low perception of operational activities by the employees, the information asymmetry among different employee groups, and that the information content is not in line with the information interests of the users. Perceptional problems are also caused by the structural and situational company factors. Bruhn ([Bruh97]) also mentions the natural limited reception, the processing capacity, and the limited memory capacity of the human beings. Picot et al. ([PiRe87], p.147) note the information overload which leads to a reduced decision quality. The higher the information overload the lower the decision quality.

5.9 Conclusion

This chapter gave us a theoretical understanding to corporate communication. This gives us the insight to define the necessary perspectives of the information space. Since it is used as an information medium to distribute and share information, it is important to know how it must be constructed to carry out corporate communication. From the latter, we derive its functions, its structures and processes and its goals. Corporate communication is important because it enables the company to act profit-oriented in its economic environment. It serves to achieve the company's objectives and improves its performance. Furthermore, it motivates the employees and increases work efficiency. It also creates the social relation in the organization and tightens the corporate culture.

We have seen, its goals are: To integrate the employees in the company activities, to reveille employee's sense of responsibility, to build trust, to satisfy employee's information needs and some more as described in the chapter. Going on, we derive that communication is organized in horizontal, vertical or diagonal. There is also a one-way and a two-way communication. The communication processes in the company are (1) between different hierarchy levels, e.g. employees communication or employee dialogue and (2) within the same hierarchy level, e.g. management interaction or employee interaction. A further important, but often neglected point is the communication competence of employees. Companies must give more attention to that factor.

[45] See research results with our research partner; a big Swiss bank. Given in chapter 1.1, p. 1.

6. Interviews

The purpose of the interviews was to gather the mindset of the enterprises concerning the Intranet, how they use it, what is important to them, and how they have integrated it in their activities. Thus, we wanted to gain an insight of the practical implementation in the industry. We therefore have conducted 20 interviews with the representatives from several enterprises. The whole interviews can be gleaned in the technical report[46] which accompanies this thesis.

This chapter first gives a comprehensive overview of the visited enterprises and the positions the interviewees hold. The second part summarizes the interviews textually and graphically and derives insights from the practical implementation in the industry.

6.1 Company Profiles

In the following, we will give a short overview of the interviewed enterprises, the number of interview partners, and their position in the company.

Industry/ Company name	Size	Activities	#Interview Partner	Position of Interview Partner
Automotive 1	C	International	1	IT Manager
Automotive 2	A	International	1	Manager Application Development
Banking 1	A	National	2	Webmaster and Member Internal Communication
Banking 2	A	National	1	Head of Distribution Management
Banking 3	A	International	1	Head Web Solutions and -Technologies
Insurance 1	C	International	1	Head E-Business
Insurance 2	A	International	1	Manager E-Business Development
Insurance 3	B	National	1	Member Corporate Communication
Insurance 4	A	International	1	Project Manager of Portal Projects
Insurance 5	C	International	1	Head Web Marketing
Logistics & Transportation 1	B	International	1	Head E-Business
Logistics & Transportation 2	B	National	2	Producer Internal Communication and Head of Operations I-net
Logistics & Transportation 3	B	International	1	CIO

[46] See [KeRi05].

Pharma 1	C	International	1	Senior Manager Business Planning
Research Laboratory	B	International	2	Communications Manager and Member Internal Communication
Retailer 1	B	National	1	Project Manager Intranet
Retailer 2	B	National	1	Head Corporate Internet Business Solutions
Telecommunication 1	C	International	1	IT Support
Telecommunication 2	A	National	1	Member Media and Public Relations
Telecommunication 3	A	National	1	Head of New Media

Table 6 Overview of the interview data

Legend:

1. To ensure anonymity of the enterprises, all names are set together of a consecutive number and the industry the enterprises are in.
2. Size is divided in three subcategories:
   A: Company has less than 5000 employees
   B: Company has between 5000 and 50000 employees
   C: Company has more than 50000 employees

Table 6 shows the industries: Banking, Insurance, Telecommunication, Logistics & Transportation, Pharmaceutical, Retailer, Automotive and Research.
We group these industries in following industrial sectors:

1. Financial Institutes (Banking, Insurance)
2. Manufacturing Industry (Industrial Manufacturer, Pharmaceutical, Automotive)
3. Service Provider (Logistics & Transportation, Retailer, Telecommunication, Research)

Table 7 shows the industries of the interviewed enterprises and their frequency in our interview portfolio:

Industry	# Companies
Automotive	2
Banking	3
Insurance	5
Logistics & Transportation	3
Pharmaceutical	1
Research Laboratories	1
Retailer	2
Telecommunication	3

Table 7 Interview portfolio

Table 8 gives an overview of the enterprises' size and their frequency in our interview portfolio:

Size	# Companies
< 5000 (A)	9
5000 – 50000 (B)	4
> 50000 (C)	7

Table 8 Company size

Table 9 gives an overview of enterprises' activities and their frequency in our interview portfolio:

Activities	# Companies
International	12
National	8

Table 9 Company activities

Table 10 gives an overview of business sectors and their frequency in our interview portfolio:

Business Sectors	# Companies
Financial Institutes	8
Service Provider	9
Manufacturing Industry	3

Table 10 Business sectors

6.2 Outcomes of the Interviews

The results of the interviews are structured in themes according to the questions posed. The main topics of the questionnaire are:

1. Function and purpose of the Intranet potentially classified by industry.
2. Change of structures and processes due to the use of the Intranet.
3. Topics and criteria for the design of the Intranet.
4. Tools and measures for quality assurance.
5. Usage of controlling mechanisms.
6. Relevance of the Intranet for the company.
7. Strengths and weaknesses of the Intranet.
8. Trend of the Intranet in the company resp. industry.
9. Where the enterprises stand concerning the Intranet.
10. What corporate goals are pursuit and how they are integrated.
11. Responsibility of the Intranet.
12. Benefits from the Intranet for enterprises and employees.

13. Conducting trainings for the Intranet.
14. What feedback do enterprises get from employees.
15. What are the most common feedback mechanisms.
16. Involvement of employees in Intranet design.

In the following, we will work out the above mentioned points in more detail, where the results are a summary of the interviews conducted. Moreover, it has to be mentioned that the purpose of the interviews was neither to see where each company stands nor to pick out an interviewee and state what she or he has said regarding a certain question. The results are also summarized in charts. It also has to be mentioned further that the interviewees were free to state multiple naming to the posed questions so that the sum of the points in the charts can exceed 20.

(1) Function and purpose of the Intranet potentially classified by industry

So far, the Intranet is used as an information medium by all enterprises. Financial institutes use it exclusively as an information medium, whereas non-financial institutes in the long run want to go further and integrate it more in their corporate activities. This means that they want to map the company's processes as far as possible on the Intranet. The trend is to transform the current information medium to a communication medium and further to an integrated work instrument, which enables information exchange, communication, and collaboration.

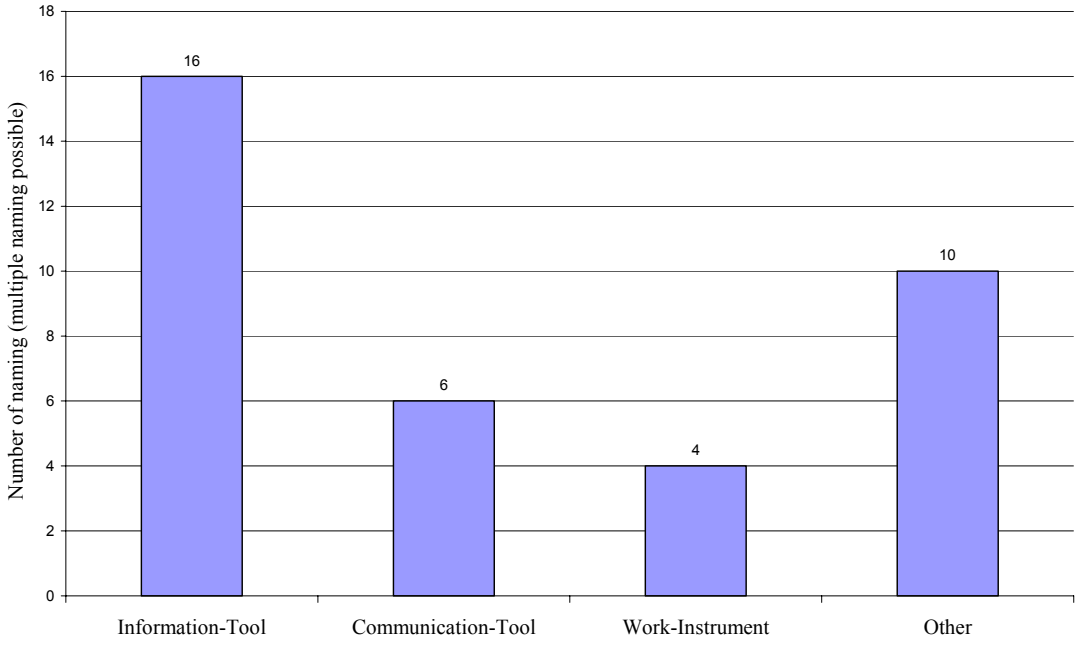


Figure 15 Functions of the Intranet

Other functions of the Intranet also have been mentioned. The Intranet is used as (1) an archive medium, (2) a document management system, (3) a knowledge sharing

platform, or (4) an e-Learning platform. The first point is a hint that in future, the Intranet will increase rapidly in size. The third point is a starting point that the Intranet will become an important knowledge sharing and creation platform in the company. The transformation of the Intranet will go from a working instrument to a knowledge management tool.
The purpose of the enterprises in using the Intranet is to support the employees in their daily business by providing them with relevant information and to increase the work efficiency.

The functions of the Intranet are displayed in figure 15. We see that the main function is the information distribution (16 naming), followed by the communication medium (6 naming), and work instrument (4 naming). In the pillar, "Other" further naming are summarized. The Intranet is also used as an archive medium, a document management system, a knowledge sharing platform, or as a support platform for the sales department.

Figure 16 shows the Intranet functions sorted by business sectors.

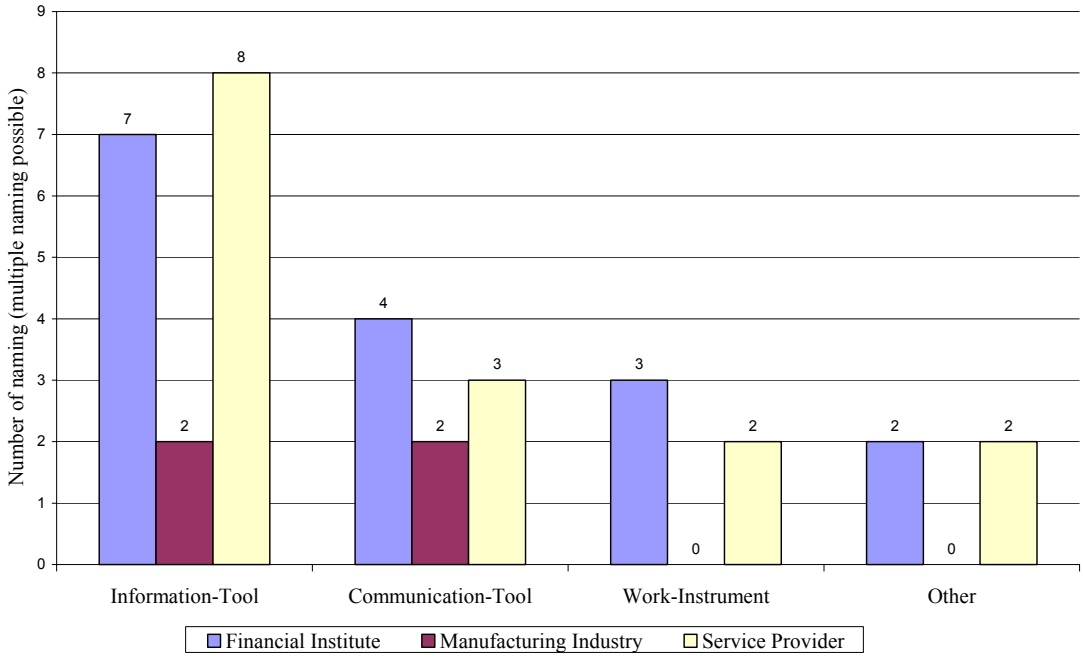


Figure 16 Function of the Intranet by business sectors

(2) Change of structures and processes due to the use of the Intranet

We have seen in the interviews that the Intranet is not changing the existing organizational structures. In contrast, the Intranet structure is aligned with the organizational structure. Based on our results, we object the statement of Hoffmann (Hoff01, p. 137) that the Intranet has impacts on the organizational structure. The organizational processes also have not been changed except for the human resources department which has been affected by the Intranet, and this is mentioned from some

enterprises. I.e. hiring, be it internal or external, is made more and more through the web. With the increased adaptation of processes onto the Intranet, the organizational processes will change. Meier ([Meie02]) states that the introduction of new communication technologies changes the work processes, lead to new interaction situations and new communication forms.[47]
The communication processes have changed due to the use of the Intranet. Information is no longer distributed by email, but put on the Intranet. It has to be specified that information is put on the Intranet if it is of a corporate-wide concern. Information which concerns only a few people is still sent by email. Also, the process of how and by whom information is put on the Intranet has changed. More quality assurance has been integrated to guarantee up-to-date information. The Intranet enables also a faster communication and annuls the information asymmetry within and between the hierarchies. It is still used as an one way communication medium. For two way communication, email is used.

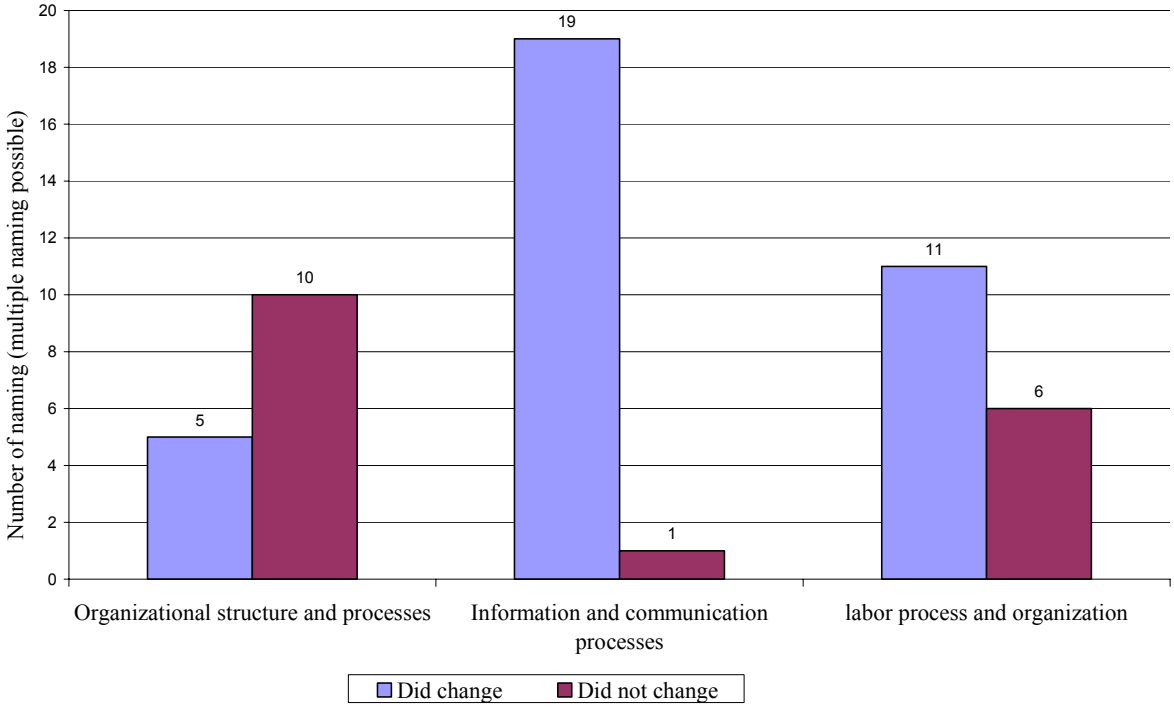


Figure 17 Organizational changes

First, we need to explain why multiple naming is possible. The reason is that sometimes only some parts of the company have been affected. E.g. only the labor organization of the human resources department may have changed, but not the rest of the company.
Deriving from figure 17, we see that in most enterprises the organizational structure and processes did not change. A statement from the interview partners was that the Intranet structure has to align with the organizational structure and not the other way

[47] See also chapter 5.5.3, p. 40, Communication technology.

round. Therefore, the organizational structure is mapped onto the Intranet. Furthermore, a not surprising conclusion is the change of information and communication processes. The one outlier appears because the IT department started to build its own Intranet satisfying its local needs. At the time of the interview, the development was still going on. The third conclusion we derive from figure 17 is that in some enterprises, there was also a change in the labor organization and processes. The company must be aware of that the more advanced the Intranet in the transformation phase is, the more the company has to take into account that not only information processes will change, but also labor organization and processes will have to be reengineered, in order to confirm to the new circumstances. Will the company neglect this fact, then a gap between information process and labor organization and process will occur, which could cause a logical break in the process chain. This break results in the loss of the advantages such as work efficiency, employee satisfaction, and so forth.

(3) Topics and criteria for the design of the Intranet

Depending on the company, the topics of the Intranet content vary understandably. The content is business oriented and must support the employees on their daily business. This is also a major requirement as discussed in corporate communication.[48] We see that the enterprises hold to this point since it is a vital factor for them. Financial Institutions have a "News" category where the most accurate news are published. This "News" category, is very important for them since information is their business. For the other industries such as Retail, Transportation, etc., news are also important, but the importance is not so highly ranked as it is for Financial Institutions. The distribution of further information is target group oriented. The Intranet is divided into areas according the target groups such as sales, human resources, category management, and management. Each group gets the necessary information, e.g. product information. Further topics of the content support the corporate structure and culture and human relations within the company. These are supported by organizational charts, team presentations, contact pages, regal affairs, travel and expenses information, ongoing and completed projects. The third topic is the provision of service such as electronic phone and address book, train time-table, a platform for selling and buying of bicycles, surfboards, theater tickets, lifts, and etc.
The structure of the Intranet is organization and topic oriented. The first developments of the Intranet were geared towards the organizational structure. Later on, further developments topic oriented structures were added so that one can navigate through the Intranet by organization or topic. For the next developments of the Intranet, two trends can be identifies among the enterprises. First, enterprises shift to have only a topic oriented navigation. Second, enterprises only have an organizational navigation because they say that since the organization is mapped on the Intranet, the employees, by surfing through the Intranet, have the opportunity to deal with the organizational structure and to become more familiar with it. There is one financial company,

[48] See chapter 5.4, p. 31.

Insurance 1, which follows another way of presenting information to the users. They divide the Intranet into six main areas which are not strictly separated, but overlap one another. The 1st area is related to the organization. The 2nd area is related to the business processes i.e. everything and every process which deals with the customer. The 3rd area contains supports processes such as human resources, IT, marketing services, etc. The 4th area contains private and business services, which are made available for the employees. The 5th area contains information about key projects of the company. Each project has its own project database, which is managed by the project manager. Finally, the 6th area contains "documentation & knowledge".

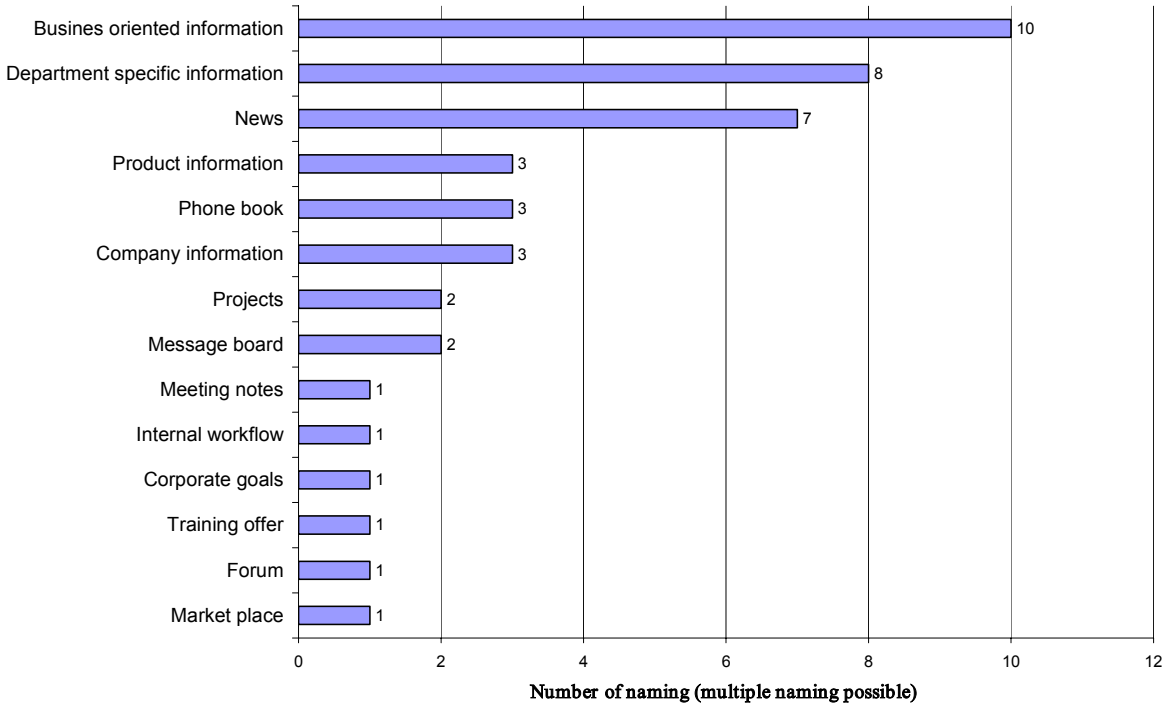


Figure 18 Topics and content of the intranets

There are several points to mention on figure 18. First, the list is not complete. Second, the number of naming is not very precise because the interviewees mentioned a few of them and finalized with "etc." Third, the most important conclusions which can be derived from figure 18 are set in the first three naming. The first conclusion is that the information provided by the Intranet must be business-oriented. The second conclusion is that each department must have its own specific area where it can publish information related to that department. The next conclusion is the importance of "news" rubric because it keeps on the one hand the employees informed and on the other hand, it ensures a high information level within the company. The second important part could be summarized as "information". This is information on products, projects, and company information. We see that the limitation between rich

and richness, discussed in corporate communication, is broken open by the new medium Intranet.[49]

(4) Tools and measures for quality assurance

In summary, it can be said that different quality assurance levels exist among the enterprises. First, most enterprises use professional content management systems (one is using front page to edit the pages) and avail the technical features these systems provide in order to assure the quality. Such features are expiry date for each page, email dispatch to the responsible person if a page expires, validity definition of pages (e.g. "this page is valid from … to …), versioning, and archiving. Furthermore, the enterprises have implemented a quality assurance process for web publishing. There are two processes:

1. Publishing news which is relevant for the entire company: Corporate communication (CC) is responsible for News publishing. The information comes either from management or is composed from the CC department itself. CC is afterwards responsible for the accuracy of the information. It decides if an update, a deletion, or a replacement is necessary.

2. Publishing information composed by different departments: Each department has the right to compose and publish information. At this level, the two or four eyes principle exists. This means that the content is looked over by the department supervisor. He can accept or refuse the article. In case of a refusal, the article goes back to the author for a revision. In case of an acceptance, the article is passed to the editor who publishes it on the Intranet. Only the editors are allowed to publish information on the Intranet. From then on, the department is responsible for the accuracy of the information. Someone of the department has to go from time to time through the pages to guarantee the quality. The company practices self-responsibility of the departments. CC goes occasionally through the web pages of the Intranet to review style, format, wording and length of the articles.

(5) Usage of controlling mechanisms

Enterprises use web tools to generate statistics such as hit rates, search words, rankings etc. These statistics are generated on a regular basis and on request of departments. Up to now, there is almost an equality between enterprises that use and enterprises that do not use further the outcome of those statistics. The results are used to improve the Intranet. An example is looking at what is accessed and what not. Based on those results, content is removed or replaced.

Although most enterprises do web statistics, we see that a lot of them do not use the outcomes to further improve the Intranet.

[49] See chapter 5.1, p. 26.

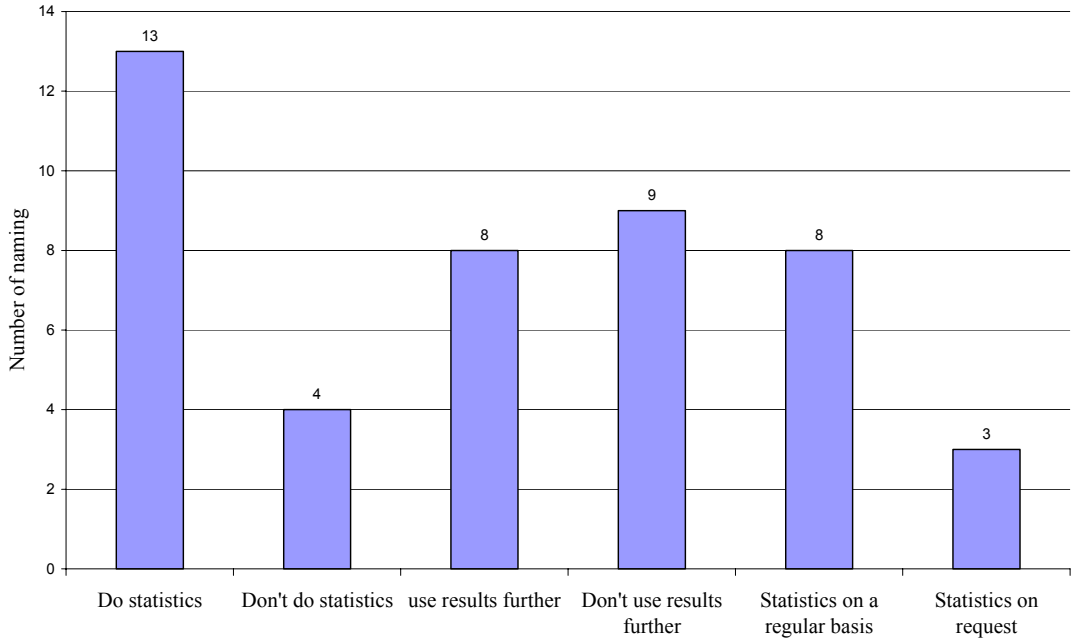


Figure 19 Usage of statistic tools

It is an equal part of those who use it and those who do not. It can also be seen that enterprises do statistics on a regular basis. Latter make statistics also on demand. There are three enterprises which usually do statistics ad-hoc.

(6) Relevance of the Intranet for the company

Every interview partner said that the Intranet is a benefit for both the company and the employees although nobody measures the benefit of the Intranet directly. "Directly" in this context means that there are no quantitative figures which indicate the success or failure of the Intranet. Information on this are gathered by employee surveys, web statistics and subjective interpretation.

Two trends concerning the importance of the Intranet can be identified. For the first group, the Intranet is not that important. The statement was that one should not overestimate its importance. The Intranet is a communication channel to provide information. There are more important applications for the company. For the second group, the Intranet is very important. One statement was also that the company could not survive for long if the Intranet would break down.

The difference between the two statements can be explained with the integration of the Intranet into the business processes. The Intranet of the second group is more integrated into the business process than the one of the first group. This means that applications and processes are integrated in the Intranet. For the second group, the Intranet is a single entry point. For the first group the Intranet is an instrument beside others. It is not a single entry point for applications or processes. There are also other ways to access applications.

(7) Strengths and weaknesses of the Intranet

The most common strengths of the Intranet mentioned by the interviewees are shown in figure 20:

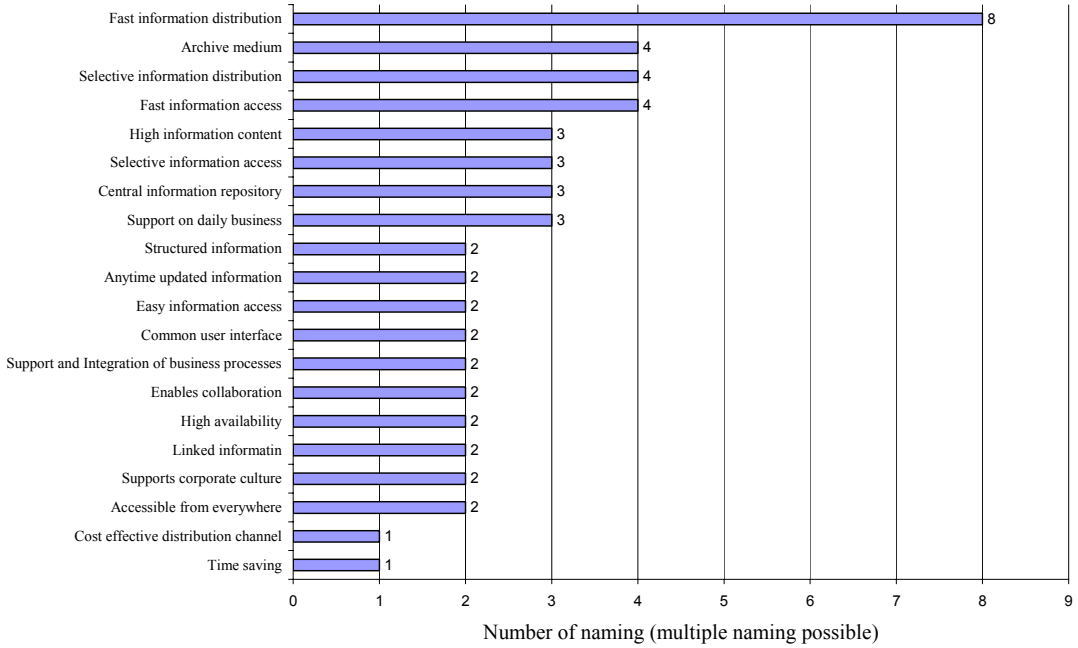


Figure 20 Intranet strengths

We see that by far the possibility to distribute information fast within the company is the most valued strength of the Intranet. This is important because it enables a fast reaction of the company to the changing environment. This is of a vital importance nowadays.[50] The second strength which tracts attention is the "archive medium" attribute. It can be concluded that the Intranet is used as a storage medium, which in turn leads to an increase of information content. If no appropriate controlling mechanism is implemented, the huge information will degenerate to an information overflow and latter to an information graveyard.

Figure 21 shows the weaknesses which were mentioned by the interviewees.

[50] See chapter 5.4, p. 32.

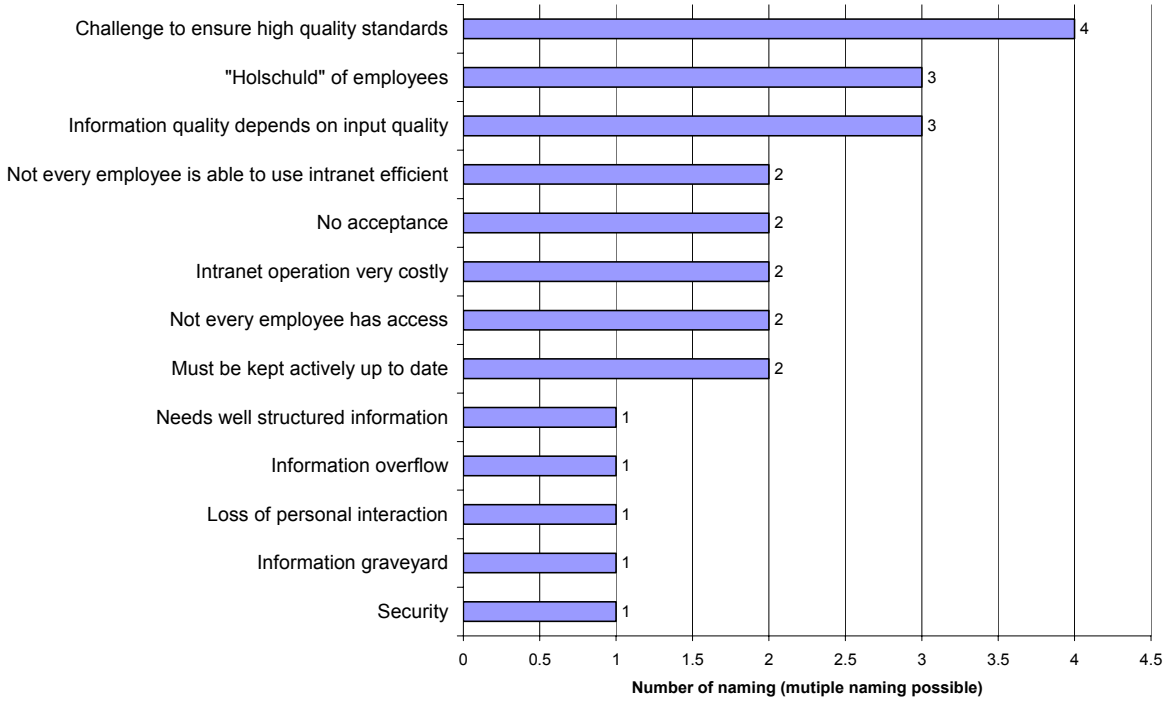


Figure 21 Intranet weaknesses

(8) Trend of the Intranet in the company resp. industry

As mentioned above, a future trend of the Intranet is the augmented integration of business applications and processes. The Intranet should become the single entry point for information, applications and processes. Some enterprises have the vision: The portal as the employee's desktop. We state that the Intranet will become a tool for knowledge creation and sharing. Figure 22 shows the transformation phases of Intranet development:

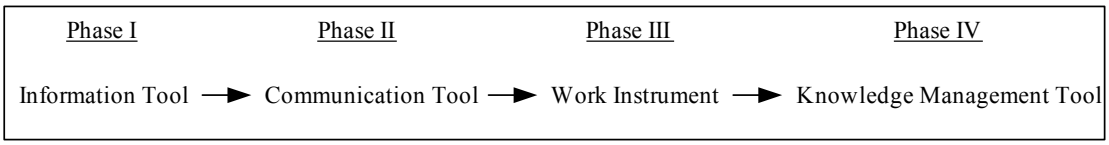


Figure 22 Transformation phases of the Intranet

As the interviewees stated, they foster the transformation of the Intranet from the information tool to become a work instrument tool. Hardly any got that far to mention knowledge management. The phases II and III (s. figure 22) also imply that the Intranet will increasingly enable collaboration between the organization members. This demands to look closer into the theories of the media usage for group work. The Media Richness Theory [DaLe86] combines the choice of the medium with the task that the organizational members have to solve together. The authors Daft et al. [DaLe86] address the issue of why organizations do process information. They state that the reasons are twofold. First, organizations want to reduce uncertainty and second, they want do reduce equivocality. Therefore, they divide the tasks in

organizations into uncertain and equivocal. An equivocal task cannot be solved even with a lot of information, whereas an uncertain can be solved if all necessary information are available. The TIP Theory[51] assumes that teams are involved in a social and organizational environment and that their behavior can only be understood in this context. The theory consists of statements concerning the nature of groups, the temporal structuring of group work, and team interaction processes. Groups fulfill three functions: (1) The production function, (2) the member assistance function and (3) the group well-being function. Group interaction processes deal with the work flow on a low level. Characteristics of the temporal structuring of group work are the chronology, the work rhythm, and the efficient assignment of time periods to actions. The Theory of Media Synchronicity [DeVa99] state that all tasks are based on two cooperation processes: Conveyance and convergence. Further, it introduces five media capabilities that are important to group work. These are: Immediacy of feedback, symbol variety, parallelism, rehearsability and reprocessability. The communication effectiveness is affected by matching the media capabilities to the needs of communication processes. DeSanctis et al. [DePo94] propose the Adaptive Structuration Theory (AST) as an approach for studying the role of advanced information technologies. AST analyses the change processes from two points: (1) The types of structures that are provided by advanced technologies, and (2) the structures that are emerged in human action as people interact with these technologies. Thus, there are structures in technology and structures in actions. AST states that the success of the media usage depends on complex, feedback-oriented mechanisms. If media usage leads to the intended results, it depends on the handling of the cooperation medium and if it fits to the requested task. The cooperation medium can be described with its structural features and its spirit. The structural features, or capabilities, are the functions of the medium, such as anonymous recording of ideas, periodic pooling of comments, or alternative voting algorithms for making group choices. The spirit is the intent with regard to values and goals of the underlying system. The spirit shows how to use and interpret the functions of the cooperative medium: How to design the decision making process? Which leadership style is propagated? Which significance has the efficiency?

Structural features and spirit build the structural potential of a system in which group are able to build their social structures. The authors introduce constructs and seven propositions of the AST.[52]

Kuppinger et al. [KuWo00] present the transformation of the Intranet to knowledge management by means of the Organizational IQ[53] method. In addition, they discuss the available technologies and the necessary processes and strategies to success.

(9) Where the enterprises stand concerning the Intranet

First, the enterprises have recognized that the Intranet is no longer considered a toy of the IT department. It is, when correctly applied, an important tool for corporate

[51] See [Schw01], p. 55. TIP stands for Time, Interaction and Performance.
[52] For further information see [DePo94], p. 127 et seqq.
[53] This method was developed at the Stanford University, [KuWo00], p. 15.

communication. The Intranet has become an important player in the corporate activities. The enterprises became aware of that it is a competition factor since it fosters fast information exchange and improves work efficiency. Therefore, the responsibility of Intranet usage has shifted from the IT to the corporate communication department. However, it must be said that competition for resources Intranet projects are second choice. Management still has to be convinced to grant resources for Intranet enhancements because the benefit, in contrast to the costs which can be showed in writing, is elusive. Nevertheless, departments having recognized its value push further the Intranet developments. During the interviews, the results came out that some enterprises plan the second or third redesign of the Intranet.

(10) What corporate goals are pursuited and how they are integrated

Corporate goals are published on the Intranet and serve as a clarification and legitimation of corporate activities. The management publishes the corporate strategy and explains and the reasons why the actions take place. The employees are informed about the situation and become more integrated into the company. This increases motivation and employee satisfaction which is an important point mentioned in corporate communication. The employees must be integrated in the short and long-term activities of the company. By doing so, the identification of the employee with the company increases and in the end, it has positive effects on corporate performance.[54] Interview partners confirmed that the employees no longer feel excluded, but more as part of the company. Schick ([Schi02]) also mentioned that corporate Intranets act as a communication channel.[55] This is also confirmed by the interviewees (s. figure 23).

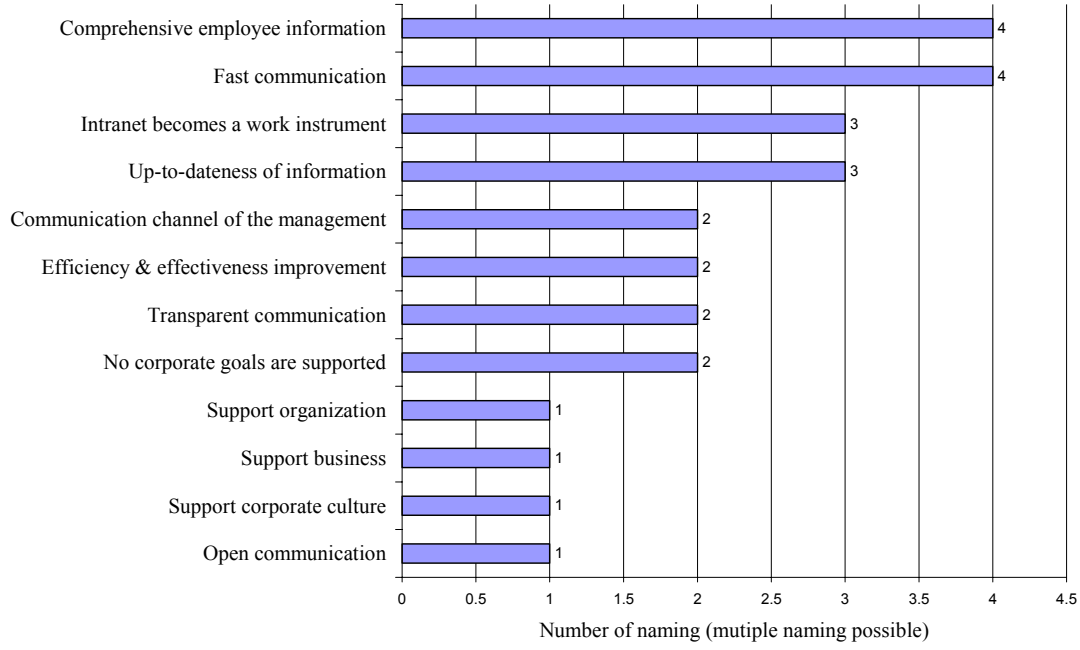


Figure 23 Corporate goals and their implementation

[54] See also chapter 5.4, p. 32.
[55] See also chapter 5.7, p. 42.

Corporate goals and how they are supported by the Intranet is shown in figure 23. Different surveys about the goals of employee and corporate communication were conducted. The results of those comply with our results. These surveys can be checked up in corporate communication.[56]

(11) Responsibility of the Intranet

It was a clear statement from the interview partners that the Intranet is affiliated to the internal communication. Only one interview partner confirmed that the Intranet is affiliated with IT. The reason is that the IT department started to build a local Intranet to support better the needs of the local workforce than those of the global Intranet of the parent company does.
According to the interview partners, at the beginning of the Internet/Intranet hype, the Intranet was affiliated with the IT department because IT started to use it as an information tool with a sort of toy character. "It was cool to have an Intranet and do fancy things with it". Over the years, the Intranet grew and people saw that it is more than just a fancy tool. They recognized that it was an efficient communication tool. People started to ask for more information and asked to be functionally integrated into it. At that time, IT realized it neither had the skills nor the requirements asked for under its responsibility. So IT handed over the responsibility of content editing and processing to internal communication. IT is now responsible for technical development and support of the Intranet and internal communication is responsible for content affairs.

(12) Benefits from the Intranet for enterprises and employees

According to the interviews, a company benefits from the Intranet. The major benefits are shown in figure 24. It can be seen that Armbrecht's view of an increased company performance can be confirmed. This statement confirms the increase in work efficiency, the cost reduction and the increase of customer satisfaction.[57] Another benefit for the company is the fast access of information. According to Mast ([Mast02]), this is important in order to secure the competitive advantage because fast communication enables fast acting.[58]

We also asked if the employees benefit from the Intranet. Our interview partners confirmed it once again. In contrast to the conclusion of the Intranet report of WEBCORP[59], employees have no benefit from the Intranet for their daily work. The contrary statements can be attributed to the different interview partners of the

[56] See also chapter 5.5.1, p. 33 et seqq.
[57] See also chapter 5.4, p. 31.
[58] See also chapter 5.4, p. 32.
[59] Corporate name has been changed by the author. WEBCORP Intranet Report 2003, „Eine Analyse Schweizer Intranets", November 2003, p. 36.

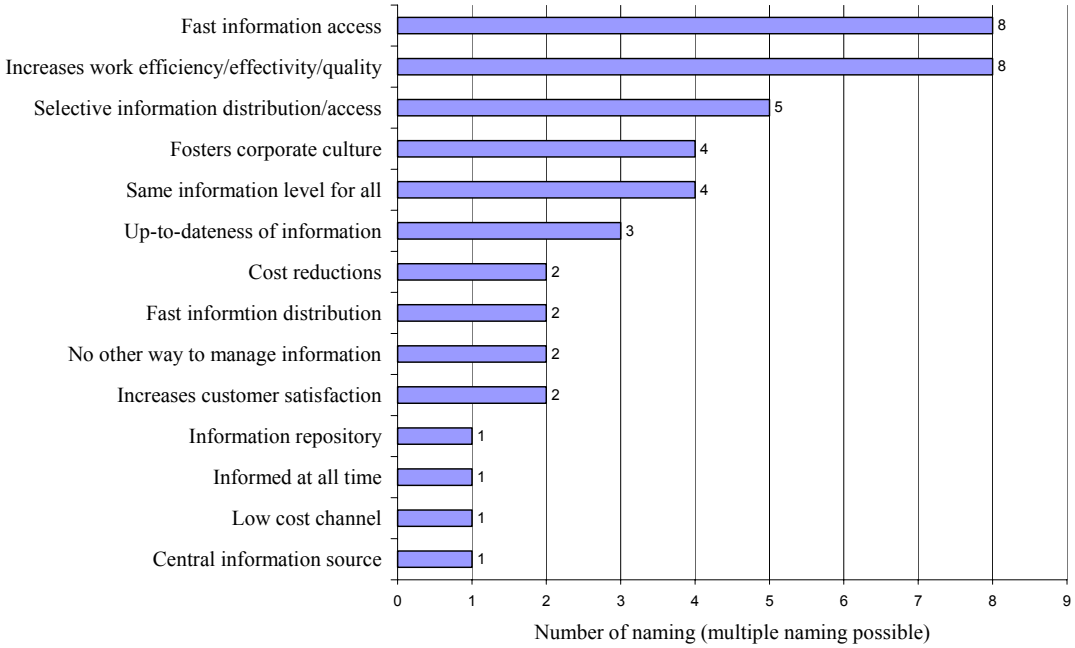


Figure 24 Corporate benefits from the Intranet

WEBCORP and our survey. WEBCORP directly asked the employees in a questionnaire, and we asked the interview partner who representatively answered on behalf of the employees. Two interview partners (Telecommunication 1 and Retailer 1) confirmed that not all employees benefit from the Intranet. The interviewee of Telecommunication 1 stated that the employees who have worked longer in the company have built their own personal relationship so that they sidestep the business processes that are mapped onto the Intranet. They know whom to call to get things done. The interviewee of Retailer 1 argued in another direction. He said that elderly employees have in some point difficulties to deal with the Intranet because they have an aversion against the Intranet. To help those employees, they conducted training sessions which in return fostered the Intranet acceptance. On a personal level, the benefit of the employee is that he also gets informed about company activities. He has the ability to look at where the direction of the company is heading towards. Also he has the possibility to get involved in the activities and get a chance to understand the action taken by the company. In turn, it results in increased employee motivation and corporate culture.[60]
The benefits for the employees are in figure 25.

[60] See also chapter 5.4, p. 32.

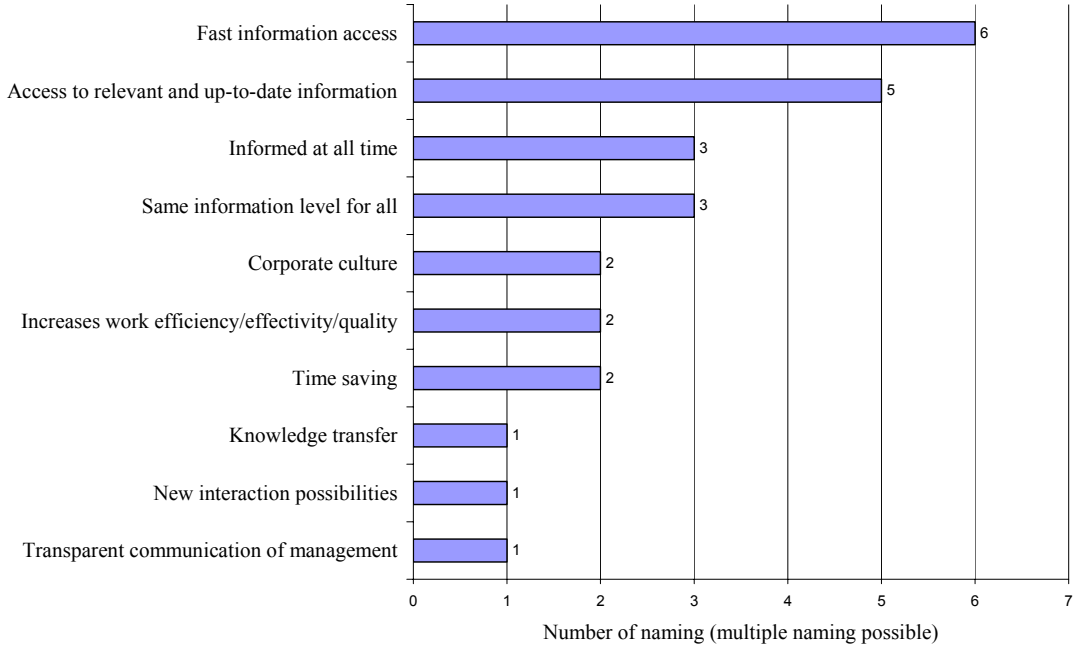


Figure 25 Employee benefits from the Intranet

(13) Conducting trainings for the Intranet

Most enterprises to not conduct any training on how to use the Intranet, although corporate communication states that it is important to train the employees in communicative and media competence.[61] Communicative competence is understood as the general faculty of speech of individuals[62] or the capability to show behavior which is not imitated[63]. Media competence is the capability of interaction with media and the understanding, perception, reception, processing and design of media content.[64] In detail, it can be said that if the Intranet is used as an information channel, then employees are not trained because the view of the enterprises is that the navigation through the Intranet should be self-explanatory. On the other hand, if workflows and web applications are integrated into the Intranet, then enterprises tend to give some training. The more workflows and web applications that are integrated, the more enterprises realize that training the employees for its use is necessary.

(14) What feedback do enterprises get from the employees

Enterprises get all sort of feedback from their employees concerning the Intranet. These feedbacks can be positive or negative, coming from satisfied or unsatisfied employees. Some enterprises get a lot, some little and some scarcely any feedback. It can be said that giving feedback or not depends on the corporate culture. One

[61] See also chapter 5.5.3, p. 39, Communicative competences.
[62] See [Hoff01], p. 70.
[63] See [Baac73], p. 260.
[64] See [Hoff01], p. 71.

interviewee stated that if there is no feedback, then everything must be alright, which stands in contrast to another interviewee who said that if there is no feedback people got used to the deficits. Another explanation which can be extracted from the interviews is that people resign because nothing happens to improve the Intranet quality after the feedback is sent. “The department does not feel responsible for this task”, was the statement.
Almost all interviews show that much attention is given to feedback. It is gathered, sorted out and used to improve the quality. People who sent feedback usually get also feedback explaining them if their suggestion can be taken into account and what will happen.
Three enterprises stated that they actively look for feedback. They contact employees directly or through surveys which they conduct once in a while and ask them about possible improvements.

(15) What are the most common feedback mechanisms

The feedback mechanisms which the employees use to send feedback are shown in figure 26.

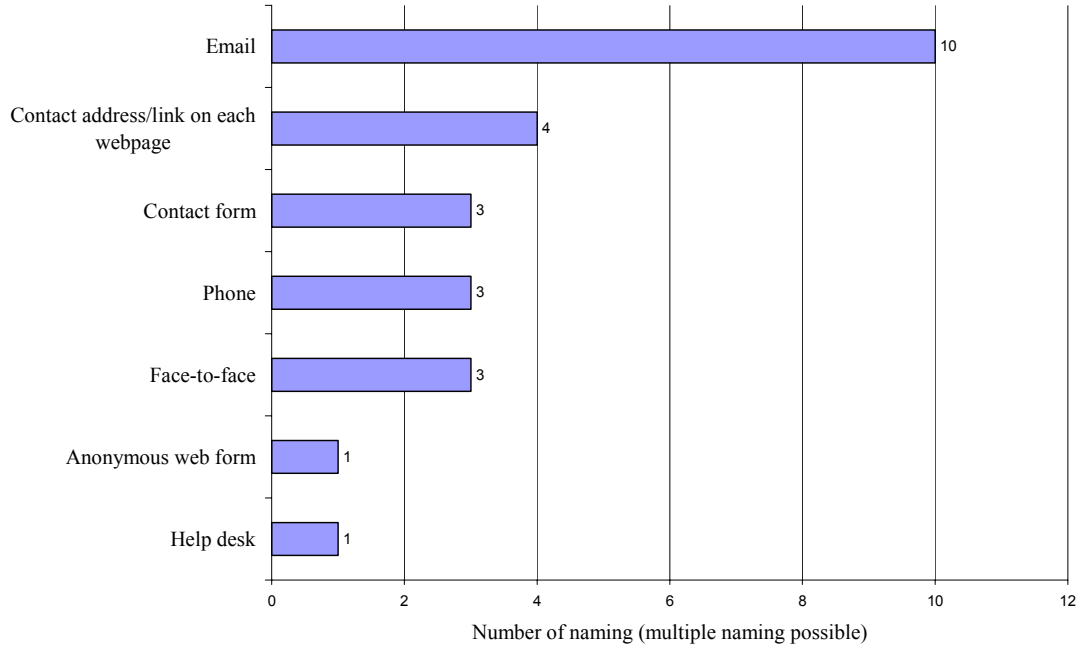


Figure 26 Feedback mechanism

We see that email is the most used tool to give feedback followed by the contact addresses and links which are on each webpage. A contact address has among others also a phone number of the person in charge. A contact link generates an email which is sent to the person in charge which is stated on the webpage. The contact form also usually generates an email, which will be sent with an except when the contact form is

a built-in form of the content management tool (CMS). In this case, the form is handled according the rules of the CMS. Hence, clicking on a contact link also generates an email, which leads to the conclusion that contact links and forms are in fact automatically generated emails.
”Face-to-face” means that feedback is given in a meeting, during a conversation or during a coffee break, everywhere where people meet and talk to each other. It is noticeable that there is one company which only uses anonymous web forms to gather feedback from employees.

Gathering feedback from employees is important because it shows if the receiver has understood the message. This is also discussed in the theoretical part of corporate communication.[65]

(16) Involvement of employees in Intranet design

All enterprises except two involve their employees in the one or another in the design of the Intranet. One of those two said enterprises have said that the Intranet was designed from the IT and human resources department.
The extent to which employees are involved in the design is restricted by corporate guidelines, standards, styles and formats. Before starting to design the Intranet resp. parts of it, the responsible project team first gathers employees, company needs, and requirements. In the interviews, different ways have been mentioned. Usually a team does not stick to one method, but applies a mix of methods to gather the requirements as accurate as possible. It can be distinguished between two main approaches. First, if meetings are conducted to gather information, then the project team invites people such as department supervisors, middle to top management, editor, authors, power user groups and division managers. To invite those people, it is important that they are able to give significant input not only from their point of view but also as representatives of the other employees. Second, when the project team wants to broaden the input by asking further the ordinary employees, they conduct surveys. A questionnaire is sent out or put on the Intranet and any employee can communicate his needs. In that way, the team gathers more input.
Usually, the project team combines both methods. In a first phase, power user group, managers, and editors are asked to give their input. After that a prototype is built, it is tested and released. In a second phase, employees are asked to use the prototype and give any input they have.
The outcomes show that the development of the Intranet is user-oriented. User’s needs and requirement must be respected when the Intranet be of any help. Those who do not deal with customer requirements will in the long run fail to generate an added value to its stakeholders. In this case, we are positive about the fact that a competitive disadvantage for the company will emerge.

[65] See chapter 5.1, p. 24 and chapter 5.2, p. 29.

6.3 Conclusion

The interviews have emphasized the following trends in internal information management:

1. A broad awareness has evolved that the Intranet is no longer a toy, yet the management is not completely convinced about the possibilities of the Intranet.
2. The fostering of the Intranet as communication medium has increased labor efficiency and employee satisfaction.
3. The enterprises increasingly map their processes to the Intranet.
4. The Intranet changes the processes and operations but not the organization structure. In contrast, it adapts to the organization structure.
5. An information democracy is emerging within many enterprises. The greater the information democracy in a company is, the harder it is for the management to lead by information hiding.[66]
6. There are several quality assuring processes which differ highly among enterprises, but there are few controlling mechanisms, if any, for the use of the medium itself.

We gained a deep insight of intranet's usage by the enterprises from the interviews. Among others, we got an understanding of its function and purpose as well as the goals of its usage, the organizational changes it implies, the controlling understanding of the enterprises, and the future trend of the intranet development, such as information tool, communication tool, work instrument and knowledge management tool.

Corporate communication and the interviews give us two different angles on corporate communication. Between both, there are differences and similarities as well as a disjoint view. This view gives us an additional comprehension about the usage of the Intranet in corporate communications. Because the Intranet as a communication channel is used for a short time period in enterprises, there is little research in literature about this topic.

A difference between the practical implementation in enterprises and the elucidations in corporate communication is that the introduction of the Intranet caused an organizational and labor process change in the company. We cannot confirm the change of the organization structure. Moreover, the interviewees told us that the Intranet structure is build according to the organization structure. These facts are associated with the medium as an information channel. Although corporate communication emphasizes the importance of communicative and media competence of employees, we saw in the interviews that hardly any company conducted training. The opinion by the majority was that the usage of the Intranet is self-explanatory and the corporate communication department assists for the composition of articles.

[66] This does not imply that leadership per se is more difficult in a company with a well-developed information democracy.

Similarities lie in the area of communication and media competences. Both corporate communication and the interviews confirm that it is important to train the employees in this area. Another similarity is that the goals enterprises pursue with the Intranet correspond with the goals corporate communication lists. Also, the increase of work efficiency and effectivity can be confirmed, as well as the motivation of employees and their better integration in company activities. The illustrations in corporate communication state that feedback is a basic element of communication. This is supported by the enterprises. They have introduced feedback mechanism in order to gather employees' feedback. Another confirmation we got from the interviews is that the Intranet brakes open the relation between rich and richness. Fast information access enables the company to act fast and adapt to the changing environment. This is illustrated in corporate communication. The interviews show that enterprises have realized it and implemented in their corporate Intranets. Enterprises have also realized that fast acting increases competitiveness. Another similarity we gain here is that corporate communication highlights that internal communication supports corporate culture. Our interviews fully confirm this statement. A phenomenon which appeared is that Intranet fosters the information democracy in the company. This phenomenon is linked to the medium and could not be forecasted by corporate communication.

The Intranet is an asynchronous channel which transmits information from the sender, e.g. management, to the receiver, e.g. employee. The problem which arises is that the employee cannot find the appropriate information. This suggests the conclusion that web controlling is neglected by the enterprises. Our study has confirmed this conclusion.

Our intention is to introduce an approach which enables web controlling and is based on the outcomes of corporate communication and our study.

7. Web Controlling Framework

The goal of this chapter is to introduce an integral framework for web controlling which is based on the insights of corporate communication and user-oriented approach. Perspectives of the information space are derived from former; latter considers user behaviour and an the ongoing improvement of the information space. We name this framework the Web Controlling Framework.

In the first part, the web controlling process is introduced and its different steps are explained. The second part introduced the necessary information space perspectives. These enable the design of the information space. Lastly we will summarize in a table, the requirements which must be solved for each perspective.

7.1 Requirements

As we have seen from our research, the most severe problems of the intranet existed in the rise of the response time of the requests, the decline of the hit precision of search requests, the impossibility to find needed information or documents, the decline of the usage of provided tools to facilitate the intranet use and dissatisfaction of the user. These problems were caused by the large number of information sources, the insufficient quality of information and service and the various existing cultures in the company. Thus, as an answer to resolve these problems, the information space must provide a high quality of information. High quality means right information, and it also implies that the information comes from the right source, at the right time, to the right consumer. Furthermore, we saw from our research with the Swiss bank that the employees complained about the unavailability of services in the intranet. This implies that a further requirement should be the securing of service availability.[67]

From the interviews, we got an insight to the publication process within enterprises. Summarized once again: The quality of information content is assured by the two or four eyes principle, where each page has a creation and expiry date and the person in charge on it, and only the editor has the rights to publish information.[68] This approach is producer oriented. The addressee is not considered in this approach. Whether he or she obtains the right information or he or she cannot find the information at all, the producer is unaware of these problems. This leads us to another requirement. We need a receiver oriented approach which means that evaluation on the receiver site must be done. Thus, it presents the user behavior.

The information space is not a static medium. Rather, it is dynamic because information, e.g. documents, directives, handbooks, is put on and removed from it on a regular basis. Another reason is that new users come along. However, what is more important is the fact that user behavior changes over time because of the learning effect of the user. The more he or she navigates through the system, the more experienced he or she gets. This asks for a cyclical approach to enable continuous improvement of the information space which is based on varying user behavior. It is

[67] See chapter 1.1, p. 1.
[68] See chapter 6.2, point 4, p. 53.

explained in previous chapters[69] that corporate communication should support corporate goals. Hence, the information space needs to assure that it also supports these corporate goals. Moreover, it should also support user's information needs which are derived from the interviews.[70]

Based on the outcomes we have gathered so far, developing a framework is essential; we name it the *Web Controlling Framework*, which must fulfill the following, discussed above and summarized below requirements:

1. Assure a high quality of service from a stakeholder perspective.
2. Provide a generic controlling process structure with the following properties:
   a. It can be applied to control the quality of the web-based Intranet.
   b. It thereby supports the achievement of corporate objectives.
   c. It is feasible from a management and financial perspective.
   d. It can be applied to control the quality of the web-based Intranet.
3. It is based on the assumption that the achievement of the quality goals can be deduced in parts from the measurement of the user interactions with the web-based Intranets.

7.2 Web Controlling Process

One problem which we face and have to solve is that the company wants to display information to the users in a useful manner, user friendly and within an acceptable time. On the other hand, the company does not know from the beginning what kind of information the user looks for. We saw from the interviews[71] that, although user's information needs where gathered in collaboration with the interviewees, the employees were not satisfied with the information provision of the company. This fact implies that despite what the organizational member says, his or her behavior in the information space does not match his subjective perception. It therefore needs an instrument with which the company is able to analyze user's behavior.

From the above two points of view, we derive the requirements so that when there is a information supply charge from the company, we name it the "Inside-out" view and where there is a behavior of the user, we name it the "Outside-in" view. In accordance to the management literature [Rüeg03] we divide the process into two parts: The "Inside-out" and the "Outside-in" view. However in contrast to the literature, we define the Inside-out and Outside-in view as follows:

> Inside-out
> *The starting point for the analysis of the information space is the needs of the company to display and present instructions and information to the user and latter's information needs itself.*

[69] See chapter 5.4, p. 31.
[70] See [KeRi05].
[71] See [KeRi05], Interview with Insurance 1, p. 36.

Outside-in

The starting point for the analysis of the information space is users' needs for information and services.

As we have defined earlier[72], web controlling is a *continuing process of defining objectives for the usage of the information space, controlling the activities, analysing the results and taking appropriate actions to reach the up-front defined objectives.* As we see from the web controlling definition, we need to define the process as cyclical in order to ensure a continuous improvement. In the following, we define the necessary actions of the process.

Therefore, we implement a process which will incorporate the above elicited requirements. We name this process the *Web Controlling Process*.

Figure 27 presents the process in which we are going to describe in this section. We start with the Inside out view:

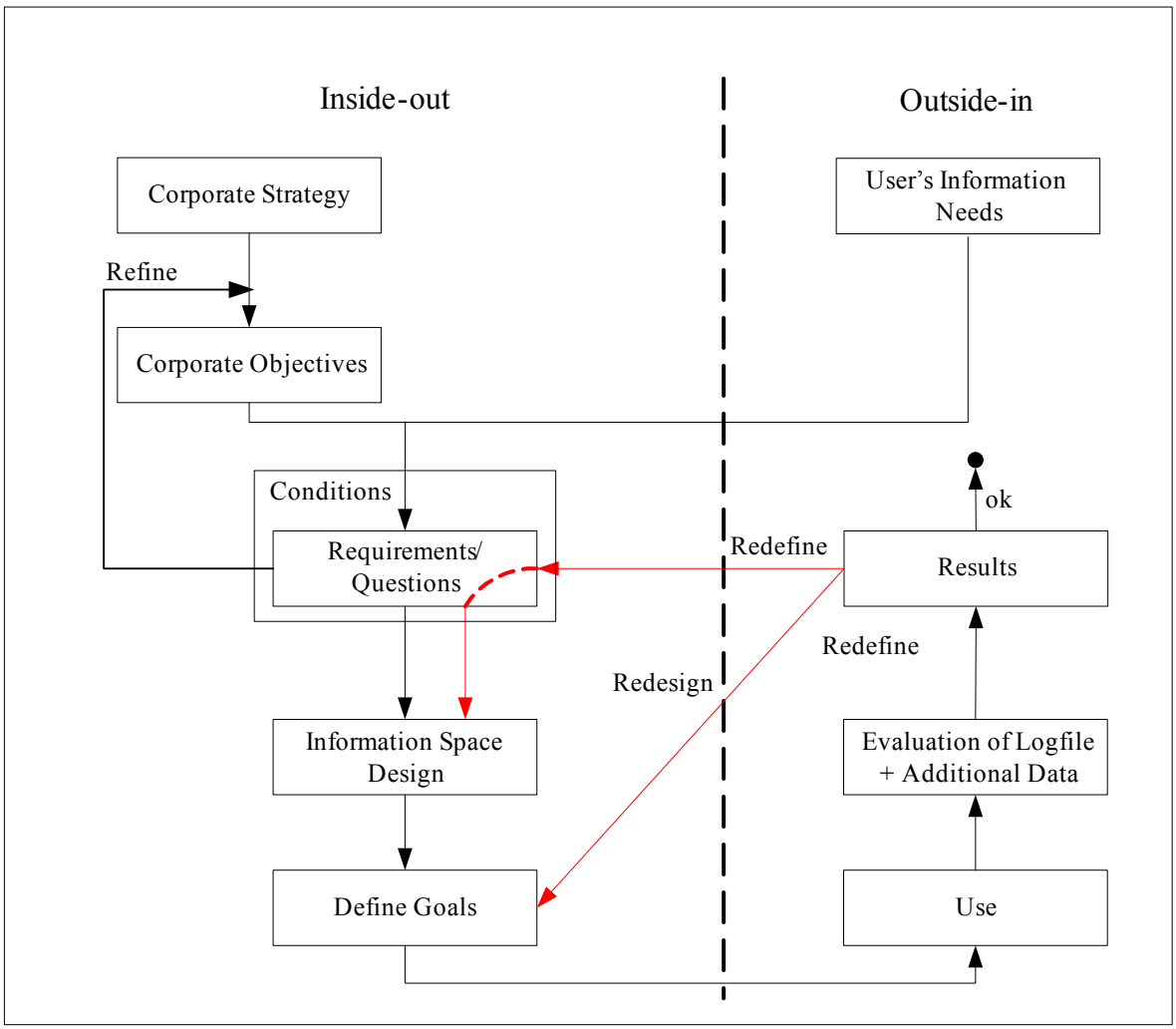


Figure 27 Web Controlling Process

[72] See chapter 4, p. 21.

From the vision of the company, the corporate strategy is derived. The corporate objectives are in turn derived from the strategy. The vision therefore sets the reference points from where the corporate strategies are derived and the goals which the company must achieve are in turn derived from those strategies ([MüLe03]). Schick emphasises that communication objectives should be derived from the corporate strategy.[73] Armbrecht states that corporate performance must be aligned with corporate strategy should it be successful.[74] This implies that corporate strategy is the starting point of our considerations.

This allows us to determine the first two steps: Corporate strategy and corporate objectives, whereas the corporate strategy is the top level from where the process starts. In the interviews, we saw that enterprises also need to take into account user's information needs.[75] The interviewees stated that thinking about which information to put on the information space, they mainly ask the organization members about their needs. This requirement sets the users' information needs on the same level as corporate strategy and goals. This means that we have two origins to start from. User's information needs is an Outside-in view because the user's point of view is taken. This viewpoint is also confirmed by Schwabe et al. ([ScKr96]). They introduce the Needs Driven Approach. It addresses the interplay between users and technology and states that the needs of persons and groups must be considered, not the exhaustion of the technology possibilities.

The strategy is defined by the top management and it indicates the direction the company will hit in the future. In the second step, the corporate objectives are derived from that strategy. If those goals are too abstract, then they have to be further refined until they become concrete. The information needs, on the other hand, are gathered from the employees by surveys or by their feedback through email, meetings or informal exchange. After defining them, we also need to take into account the conditions of the company. These are the rules of corporate communication which set the borders for the corporate activities.[76]
The next step in the process is to make sure that the information space is able to fulfill the identified goals. In doing so, we need to identify and define the necessary requirements resp. questions, which the company must resolve in order to achieve the goals. The requirements and questions are mapped to the quadrant II of the spiral model.[77] Afterwards, the information space will be design according the requirements
In order to avoid confusion it must be said that the requirements we identified in chapter 7.1 are the demands on the web controlling process. The requirements resp. questions we talk about in this paragraph are the demands on the information space.

The fourth step of the process is the actual design of the information space according the requirements. The last step of the Inside-out part is the definition of

[73] See chapter 5.7, p. 42, [Schi02].
[74] See chapter 5.4, p. 31.
[75] See Interviews in [KeRi05].
[76] Conditions and rules are defined by corporate identity, see [KeRi05].
[77] See chapter 8, p. 76 et seqq. The Spiral Model.

goals[78] which allow us to review the success of our design. I.e. the success from results of step 9 of the controlling process has to match with the targeted goals. These goals are the entry point into the spiral model.[79] They describe if we have succeed in what we intended to do. The following example should provide clarity. Let us assume that a corporate objective is to provide the corporate strategy to the employees. The goal which can be defined is: *We (the management) would like to see that 85 percent of our staff will have a look at the strategy.* By defining this goal, the management is now able to see whether the aim has been achieved or not. If not, then the management has to ask itself why the goal was not achieved. Perhaps the information was "hidden" and must be replaced and so forth.[80] After the last step of the Inside-out part, the information space is going live.

In this section, we are going to describe the Outside-in view: The stakeholders are able to use it and through their use, they will generate log file data. These data will be evaluated and some results will be generated. So, the next steps of the process are the usage of the information space and the evaluation of the log file data. These data represent the user behavior.

The comparison of the results with the targeted goals will show whether they have been achieved or not. If they have been achieved then there is no need for further improvements of the information space. If not, then there are two possible ways to redefine the objectives:

1. Redefine the requirements.
2. Redefine the goals.

After that, the process is continued as described above. The last part of the process, from the "use" to the "results", is the Outside-in view. Here, we see how the users behave in the information space and if the latter really displays the information and services the former need. It carries the view from the "outside world" into the system. This Outside-in view represents the quadrants III, IV and V of the spiral model.[81] These quadrants define the evaluation process to extract user behavior.

The web controlling process can be seen as a Boehmian spiral which permits continuous improvement.

7.3 The Web Perspective Schema

The goal of the web perspective schema is to determine the requirements resp. questions mentioned in the web controlling process. The schema has six perspectives to determine the requirements (s. figure 28). In the following, we will point out how we achieved these perspectives.

[78] For an example of such goals see chapter 8, p. 79 et seqq.
[79] See chapter 8, p. 77.
[80] As an example see chapter 9.6, p. 107. Maybe the view of the user is distracted.
[81] See chapter 8, p. 77.

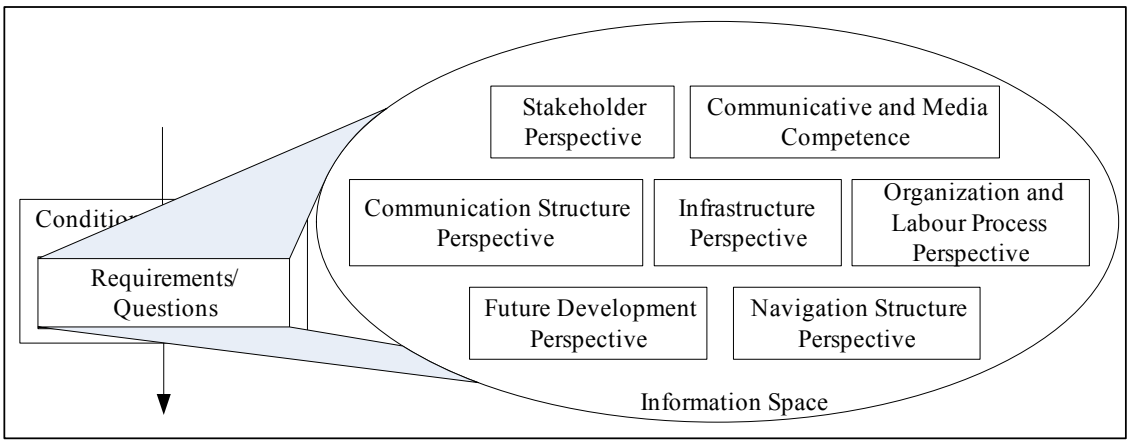


Figure 28 The Web Perspective Schema

Stakeholder Perspective (roles of the participants)
The stakeholder perspective is divided into two parts. The first part is the *lead stakeholder*. He is the stakeholder who initiates the creation of the information space (e.g. manager, employee). The target stakeholder indicates who will be affected by the implementation of the information space (e.g. employee, customer).[82]

Communicative competence and Media Competence Perspective
As we have seen from the organizational communication[83] and the outcomes of the interviews, it is important that the organization members have the competence to deal with the information space. On the one hand, they need communicative competences to communicate verbally and non-verbally with each other. On the other hand, they need media competences to interact with the information space and to understand, to absorb, to process, and to design media content. In order to foster these competences, the company must set up training courses to cover employee's deficits.

Communication Structure Perspective
This perspective deals with the information and communication structures, and information and communication processes of the company. It determines the communication directions, the communication purpose, the communication level, and the communication flow. The perspective encourages thinking about the changes and redesigns of the communication processes caused by the new communication channel.[84]

Infrastructure Perspective
The infrastructure defines the technology that will be used in the information space and the services that will be provided to the users. This is derived from the requirement[85] to provide a robust infrastructure and a high availability of services.

[82] Derived from chapter 5.2, p. 28 et seq.
[83] See chapter 5.5.3, p. 39 and chapter 6.2, point 13, p. 61 and interviews in [KeRi05].
[84] See chapter 5.5.2, p. 36 et seqq. and chapter 6.2, point 2, p. 49.
[85] See chapter 7.1, requirement 1, p. 66.

Organization and Labour Process Perspective

The process perspective sets the procedures and processes to support the advance of the information space. The more the information space moves rightwards in the Intranet transformation phase[86], the more sophisticated the information space gets.

The perspective defines which processes of the company to put on the information space, how the existing processes must be changed or re-defined in order to support the corporate objectives. The perspective does not stop at the information space, but it rather has to be analyzed which other company-wide processes are affected.[87]

Furthermore, the introduction of the information space might also cause changes in the labor process.[88] Due to the introduction of the information space as an additional communication channel labor processes must be aligned. The organizational responsibilities must be defined to assure a high quality process. E.g. only the editor is allowed to publish information or the content must be reviewed from two department members. This is also called the four eyes principle.

Navigation Structure Perspective

This perspective defines how the navigation structure of the information space is designed. First developments of the information space were aligned to the organization structure of the company. Later on, topic oriented structures were added so that one can navigate through the Intranet by organization or topic. New approaches divide the information space into different, but overlapping business areas.[89] The navigation structures are, but not limited to: *Organization-oriented*, *topic-oriented* or *business area oriented*. These perspectives can be used individually or be combined with each other. In the future, more navigation structures can emerge.

Future Development Perspective

It sets further perspectives on how the information space will be used in the future and what goals are to be achieved. The transformation phases of the information space are:[90] Information tool → communication tool → work instrument → knowledge management tool. In this context, the management has to decide at which phase the information space is at the moment and wheter the information space should be transformed further. The different implications of media usage must be taken into account during the rightwards movement of the information space.[91]

[86] See transformation phases, p. 56.
[87] Derived from the interviews, chapter 6.2, p. 49 et seq.
[88] E.g. see [KeRi05], Interview with Insurance 5, question 6, p. 54.
[89] This perspective is derived from chapter 6.2, point 3, p. 51.
[90] See transformation phases, p. 56.
[91] See theories of media usage, p. 56 et seq.

Example

The following example shall show how the process works.

From the corporate strategy, we derive the corporate goal "*to improve the internal corporate communication through the use of the Intranet.*" Since the goal is too general, we refine it further. "*Management directives and decisions must be communicated in the future through the information space.*" This is concrete enough to continue to defining the questions we need to resolve. We use the perspective schema in order to generate the questions.

Lead stakeholder perspective: The lead stakeholder is the management and its goal is to communicate the directives and company decisions through the information space. It is mass communication because all employees must be informed.

Target stakeholder perspective: The target stakeholder is the employee of the company. His goal is to get the information and orders fast, and in a useful and understandable manner.

The infrastructure perspective: Defines which technology to use. In our case, we use open standards (Java, J2EE) for the implementation of the information space. We also need to define which services to provide. A good one would be a push service, which feeds the user with information on his personalized information space page. The employee can subscribe for this service. Thus, he gets a notification whenever a new instruction is put on the information space.

The organizational process perspective: In order to support the overall corporate goal, we need to change the communication processes within the company, which means that, "*every directive and decision taken by the management must now be communicated through the information space. Bulletin boards are no longer allowed.*"

The communication structure: The purpose of the goal is to inform the employees. It is a one-way, vertical communication, meaning that the management informs top-down the organizational members and the members are not given the possibility to communicate with the management. Furthermore, it is an asynchronous communication and single-level. It is the latter because the instruction goes directly to the employees without having a hierarchical level in-between process the information further.
The navigation structure: Deals with the navigation within the information space. One has to think about how the navigation has to be designed. The development could correspond to the organizational structure, or to a topic oriented structures. Possible is also the development of a business area oriented structure. The mentioned navigation structures are derived from the interviews.[92] Other navigations are also possible.

[92] See chapter 6.2, point 3, p. 51.

The development perspective: Deals with the information spaces' future asking about the role that the information space will take in the future. In our case, the management says that "*the information space will take an important role within the corporate communication. In the future, any communication to the employees will be made through the information space. Therefore, we need to further expand the communication and put more resources on it. In the next three year 70% of our communication shall be made through the information space. We achieve this goal by educating management and employee and further align the communication processes to it.*"

7.4 The Web Perspective Framework

The framework presents the summary of the chapters discussed before. Table 11 shows which requirements must be solved in each perspective in order to adapt the information space to the user's information needs or the corporate goals.

Perspectives	Requirement to solve/ Questions to answer
Stakeholder	1. Define lead stakeholder 2. Define target stakeholder 3. Define type of communication: 3.1 Interpersonal communication 3.2 Organizational communication 3.3 Mass communication
Communicative Competence	Foster communicative competence and provide training
Media Competence	Foster media competence and provide training
Communication Structure	1. Define communication structure: 1.1 Purpose of communication 1.2 Vertical, horizontal, diagonal communication 1.3 One-way or two-way flow communication 1.4 Synchronous – asynchronous communication 1.5 Single-level – multi-level communication
Infrastructure	1. Define the technology to be used 2. Define the services needed
Organization and Labor Process	1. Define the processes to be put on the information space 2. Change processes if necessary 3. Put processes on the information space 4. Determine which other processes are affected 5. Define organizational responsibilities for different actions in the processes 6. Take appropriate actions for a logical continuation of the processes
Navigation Structure	1. Define navigation structure of the information space (e.g. organization oriented, topic oriented, business area oriented)

Future Development	2. Define the current transformation phase of the information space 3. Determine the future transformation phase of the information space 4. Consider the principles of media usage 5. Reconsider the Perspective Framework

Table 11 The Web Perspective Framework

7.5 Conclusion

The web controlling framework facilitates the implementation of web controlling. The framework is divided into two parts. The first part is the Inside-out approach in which we looked into in detail. This part shows which information space views must be considered in order to meet corporate objectives and user needs for information.
Not until having the information space deployed and given free for use, we can see how it is really used. And here the second part of the Web Controlling Framework appears, the Outside-in approach, which has an important role. This part of the Web Controlling Framework analyses the usage of the information space by its users. It indicates user behavior, documents accessed, surf paths and so forth. This is done by means of the analysis of the web logfiles. It is not enough just to define the information space and believe that the user's needs will be met. As an interview partner mentioned; "*Although we discussed thoroughly with our people their needs and requirements, we got some nasty feedback from the users telling us that they cannot find the needed information, that the structure is not user-friendly and of likewise.*"[93] This example shows that it is not enough to count only on the input of the users. We need to see how the users really navigate in the information space.

[93] See [KeRi05], Interview with Insurance 1, p. 36.

8. The Spiral Model

This chapter introduces the spiral model. It is the concretion of the Outside-in view of the web controlling process. The steps of this view are mapped to the spiral model and implemented as an evaluation process.

The spiral model was first introduced by Boehm ([Boeh86]) as a procedure model for software development. We use the concept of the spiral, but alter the model to satisfy our specific needs. In contrast to Boehm's spiral which is divided into four parts, we divide it into five parts. The aim is to use this model in order to describe a process on how to analyse and evaluate the information space or parts of it. In the following, the model is described. Afterwards, some examples are given for a better understanding of the application method of the spiral model.

8.1 Introduction to the Spiral Model

The following section explains the build up of the spiral model and the different steps in the controlling process:

First, in quadrant I, we identify goals which must be satisfied by the system. This complies with the definition of goals in the web controlling process.[94] In quadrant II, we plan the appropriate actions to be taken in order to achieve the defined goals. This corresponds to step 3 in the web controlling process. The actions which must be taken, comply with the web perspective schema.[95] That means that the different perspectives must be thought over and mapped accordingly to the perspectives. In the web controlling process, the perspectives have already been implemented so why look at them at this point again? The reason is due to the fact that the definition of goals in some perspectives must be aligning to the goals. An important point in this quadrant is the determination of the measurement method. Questions which arise here are: (1) How to measure the goals and (2) what is the threshold which indicates that the goals are achieved. This is important as the generic filter pipeline[96] which can be completed with different components to comply to the measurement method. In quadrant III, we finalize the implementation of the perspectives and implement the measurement component. Quadrant IV describes the usage of the system resp. the information space by the users. In this phase, the system gathers information through the generation of web log files about its usage. Quadrant V describes the evaluation of the data and validation of the results. The evaluation is made with the generic filter pipeline.[97] The evaluation of the data complies to step 7, evaluation of log files of the web controlling process. At the end, the results are analyzed and compared with the goals defined in quadrant I. If the goals are satisfied, then no further actions must be

[94] See chapter 7.2, p. 68.
[95] See chapter 7.3, p. 71.
[96] See chapter 10, p. 111 et seqq.
[97] See chapter 10, p. 118.

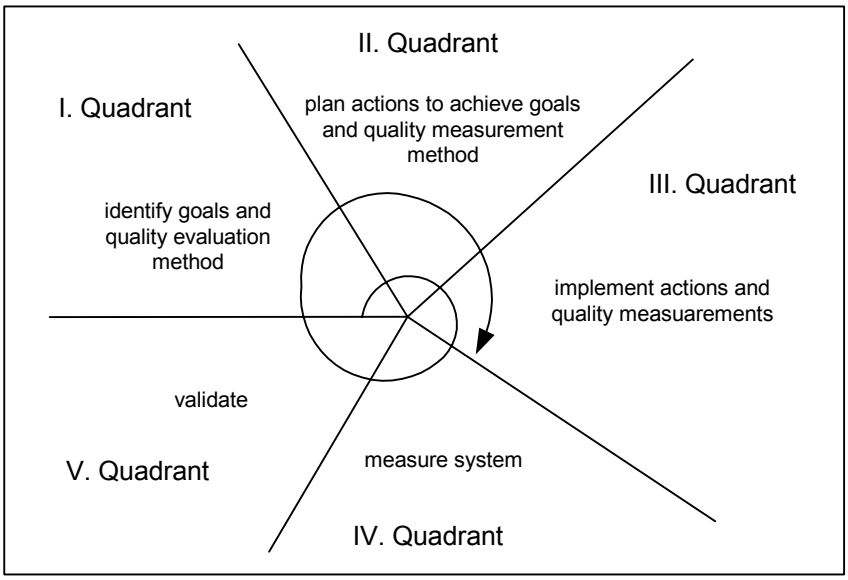


Figure 29 Spiral model

taken. If the results differ from the goals, then further appropriate actions must be implemented. These actions should change the behaviour of the system in favour of the goals. Step five implements the actions. The last step corresponds to the last step in the web controlling process. After the entire cycle is ran through, the process can start again from the beginning.

A strength of the spiral model is that not only the entire information space can be analyzed, but also parts of it can be analyzed. Let us take the infrastructure perspective[98] for an example. In this perspective, web services can be defined to support the users in their work such as who's who, electronic phonebook or search agents. The spiral model gives now the possibility to analyzed one of these services. Questions like, "how heavy is the electronic phonebook used?", "how is the search agent used?", and "does the search agents cause any benefit?" can be examined.

The web controlling process is used on a macro level. The macro level is the view on the entire information space. I.e. it is looking from outside to the information space as a whole. The spiral model is a controlling cycle containing the four parts: Plan – Do – Check – Act (s. figure 30). This controlling cycle can be used on a macro level or on a micro level. Used on a macro level, it analyses the information space as a whole and verifies whether different perspectives have been accurately implemented. On a micro level, the spiral model is able to analyse parts of the information space as mentioned above with the service who's who or search agent. This process will be explained in more detail with the examples below.

[98] See chapter 7.3, p. 71.

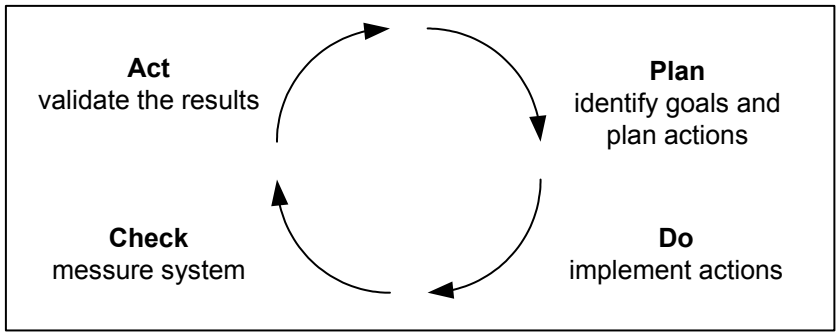


Figure 30 Controlling cycle

We introduce some examples and show how the spiral works:

Example 1: A Swiss bank introduced a stationary information agent to help the employees find information on the Intranet. The agent should assist the employee on the information retrieval. The user could enter his preferences like *title, author, topic etc.,* and the agent would search on the Intranet for papers, documents and other information containing those keywords. The agent was also able to run offline, which means that it would look for information during the night and put the results into the user's mailbox. It could also run periodically, looking for information once a week or every other day.
We have the task to analyse how useful and efficient the agent works.

At first, we define the system to be analysed. Our system is the information agent without the interactions to third parties (s. figure 31).

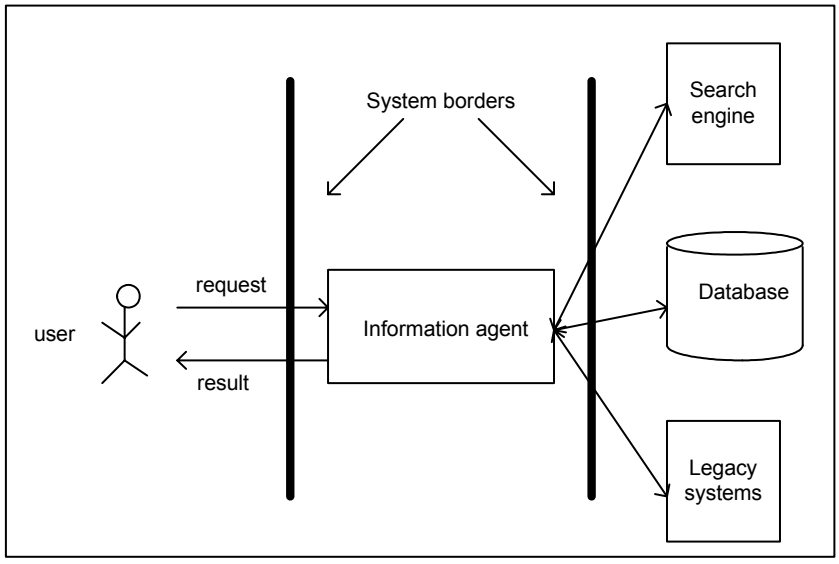
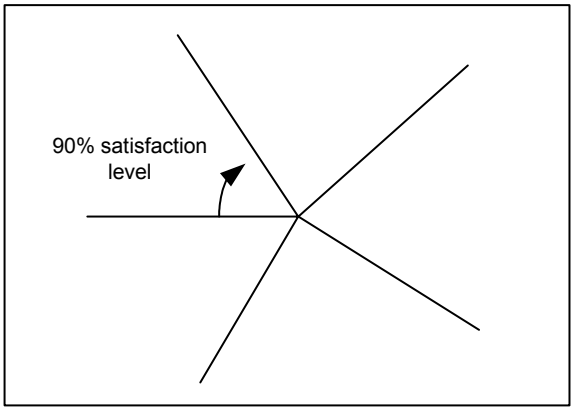


Figure 31 Information agent

Step 1: Identify goals. In this step we formulate our expectations. Furthermore, we also think about a quality evaluation system on how we test the system and what our quality standards are.

We like to see that the requests from the users are satisfied 90 percent or more. In other words, 90 percent or more of the results satisfy the expectations of the users.

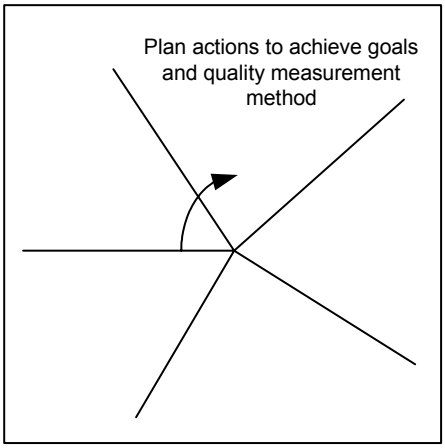


The goal is set into the first quadrant.

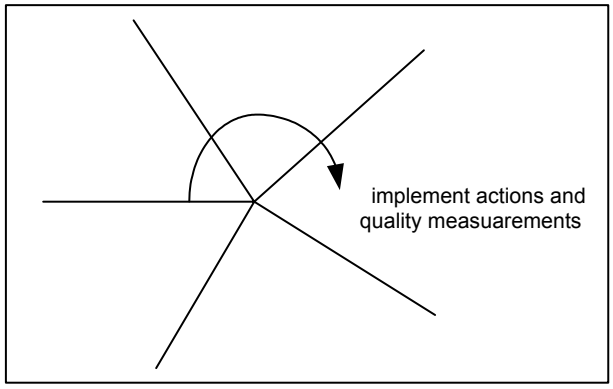
Step 2: In this step, we plan the actions which have to be taken in order to achieve the goals.

1. *Improve or rebuild information agent system.*
2. *Educate users to utilize the agent system better.*

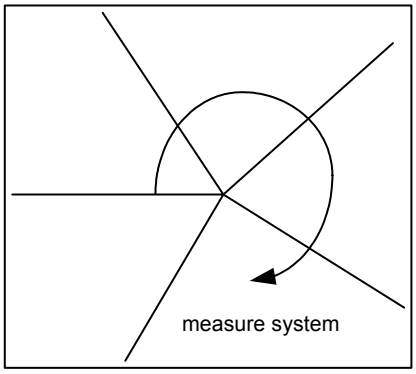
We test the satisfaction of the users with a survey. We create a questionnaire and hand it out to the users. We gather the questionnaire which are returned to us and generate our result. We are aware that the quality of the results depends on the rate of return, the quality of questions, and answers. We also think that only the users can give us a feedback on how satisfied, since only they know if the information they get is what they have looked for.



Step 3. We implement the actions we have defined in step 2. We also hand out our questionnaire.



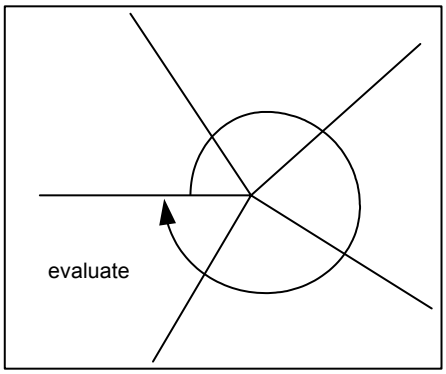
Step 4: We request the system to gather relevant output data. We gather the returned questionnaire.



Step 5: We compare the results with our initial goals.

Do the results satisfy the 90 percent rules?

To answer this question, we evaluate the questionnaire and see if the users are satisfied with the agent's proposal in 90 percent or more of the cases.



Example 2: Marketing website

A job agency wants to analyze its website to see if it meets its goals. The motivation of the agency to design a website and put it on the web is, on the one hand, customer retention and loyalty and on the other hand, customer acquisition. Therefore, we have to answer the question on how we retain or acquire customer. First, we identify our goals and formulate questions. The questions to research might be:

- *Did we achieve our expectations regarding user registration?*
- *How many registered users actually do contact us?*
- *How many new customers do we acquire per period?*
- *How many users revisit our website?*

Answering those questions can help us reach a solution which resolves our needs.
Furthermore, we have to think about how we measure our success. What is our measuring unit, our leveling board?

Step 1: According to our spiral model, we first identify our goals and the quality measure.
Our goal is that 15 percent of the users will register to our website.
How will we measure this? Our approach is to look at the web logfiles and extract the appropriate information. We will describe this approach below.

Step 2: We now identify the actions we have to take in order to achieve our goals and define our measurement method, too.

- *Design or re-design website regarding:*
  Design standards, user friendliness, navigation behavior.
- *Introduce Direct-Marketing*

Our measurement method is as follows: First, we extract our data set from the web logfile. The logfile includes all the requests which have been made to the website. The web logfile requests look like following:

555.212.34.76 - - [08/May/2001:14:17:56 +0200] "GET /products/info/index.html HTTP/1.1" 200 16788 "" "Mozilla/4.0 (compatible; MSIE 5.5; Windows 98)"

The data set is all of the logfile or a part of it. After that, we group requests into sessions. A session is a set of requests from the same origin and with a maximum time difference of Δt (Δt variable and therefore can be set by the researcher) from $request_n$ to $request_{n+1}$ ($1 \leq n \leq N$, N: set of requests). We count the sessions which contain

a "register"-request. The "register"-request in the web logfile would look like the one below:

555.212.34.76 - - [08/May/2001:14:25:26 +0200] "POST /cgi-bin/register.perl HTTP/1.1" 200 16788 "" "Mozilla/4.0 (compatible; MSIE 5.5; Windows 98)",

and we set these sessions in relation to all sessions which have been derived from our data set. If the ratio is 15 percent or higher, then we have achieved our goal. If it is lower, then we did not complete our goal.

We have to take into consideration that the origin of the user might be invisible. This happens if the requests go through a proxy server or a firewall ([Mato00]). The proxy or firewall hides the original IP address of the user and displays its own instead. Techniques to identify distinct users through proxies and firewalls are through the use of cookies, hidden fields, and unique URLs. In our work, we will not go further into details of those problems. We assume that we have distinct users.
Another problem as described in [Mato00] is browser caching. This technique is used to speed up the response time of requests. A visited webpage is cached in the browser. A new request of the user to the same page will reload the cached page. The request will not go the whole way to the webserver. A way to resolve this is to introduce the HTTP order([Mato00]) "*Pragma = no-cache*" into the webpages. We can now force the browser to request the webserver.
Last but not least, we also have to consider multiple requests from the same user. A user can visit a site many times. In this case, we must count the user only once otherwise we would falsify our results. Let us illustrate it with an example.
A site has ten requests. Therefore, we assume that we have ten users. If one out of this ten is a registration request then we have a registration ratio of one tenth. This means that 10 percent of the users do register.
But if three requests are made by the same user, we actually have only 8 different users. Now, the registration ration is one eighth, which is only 12.5 percent of the users register in this case.

Step 3: We implement the actions
- Design the website –
and also set quality standards for the implementation. These standards depend on the actions we define and vary accordingly.

Step 4: Request system
Put the website on the web and gather information from the web logfile about user and registration behaviour.

Step 5: Evaluate results

We parse the web logfile, read the data, and generate our results.
Did we reached our goal concerning the 15 percentage limit?
If we did not reach our goal, then we have to identify either new ones or alter the old ones. We have also to think about our actions. Perhaps they are not appropriate for our goals. We might have considered the content rather than design standards and user friendliness.

Example 3: Publishing on the Intranet

A big Swiss bank wants to introduce a content management tool to facilitate the publishing process. The reason why the bank wants to facilitate the publishing process is that it wants to improve the speed of displaying web content and reduce the error rate of old or not valid content on the Intranet. We would like to analyse the goals and expectations of the introduction of this task.

Step 1: .Our goals are as follows:

- *We want to reduce the errors of having old or not valid information on the Intranet by 80 percent.*

The more difficult question is how we will measure this goal. One way is to do a survey with a questionnaire. This will not give us precise results because the results would very much depend on the rate of return and the quality of the questions and answers. Furthermore, there is also a range for interpretations which could alter the results further. A second way would be to implement a tool, a spider, which periodically parses the Intranet and archives the websites. It then generates the Δ between the old archived web and the new one. The problem we face here is how the spider recognizes old and invalid information on the web. Here, we get into technical problems.

Step 2: *We introduce a content management tool which has several functions and parameters which allow us to parameterize our process.*
Functions and parameters should contain:

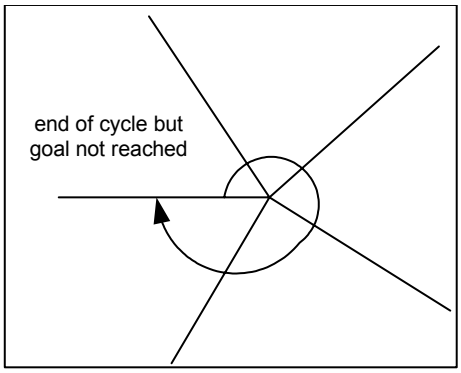
- *Author, creation date, validity date, display content, remove content, display to certain interest groups, mark as valid/invalid.*

We decide to create a questionnaire and hand it out because we believe that tool would be too resource-intensive and cost-intensive. We also have the opinion that the questionnaire survey will give us good enough results.

Step 3: *We buy and if necessary, we customize the tool. After that, we integrate it into our Intranet.*
The users must use the new tool for publishing on the web.

Step 4: *Since the users work with the new tool, we conduct the survey and gather information about the errors. These include valid and invalid information, and old information published on the web. We have to be aware that these are subjective impressions of the users.*

Step 5: *We evaluate the errors and compare the result to our goal. Could we decrease the errors by 80 percent with the new tool?*
Looking at the result, we see that we could not decrease the errors by 80 percent. We did one cycle of the spiral but unfortunately we could not achieve our goal. What are we going to do next?



Our next step is to do a second round on the cycle and redefine some of our approaches. We have different ways to go and redefine them. One thought is to say that the content management tool is not good enough and that we have to change it. Another suggestion is to say that the results are not precise enough. The questionnaire did not work as we thought. We decide that we need more precise results to make a clear statement and therefore, we implement a spider to gather more precise results.

We reenter the cycle.

Step 1: Our goal:
- *We want to reduce the errors of having old or not valid information on the Intranet by 80 percent.*

There is no change to our initial goal.

Step 2: *We introduce a content management tool which has several functions and parameters which allow us to parameterize our process.*
Functions and parameters should contain:

- *Author, creation date, validity date, display content, remove content, display to certain interest groups, mark as valid/invalid.*

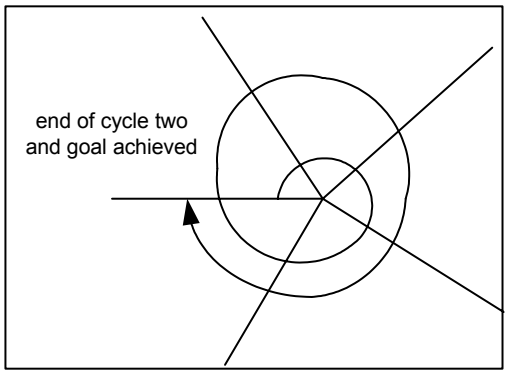
We decide to implement a spider. We do the differentiation of old and new information with the aid of a creation date. The spider recognizes old information if the creation date is older than a given date.

The problem we face here is what are we doing with the information which is already on the web? Either we sort out the old information by hand and start using the CMS from then on or we re-implement the website with the CMS tool.

Step 3: *We buy and if necessary, we customize the tool. After that, we integrate it into our Intranet.*
The users must use the new tool for publishing on the web.
There is no change to our initial goal.

Step 4: *Since the users work with the new tool, the spider is able to archive the site and generate results. These include valid and invalid information, old information published on the web.*

Step 5: *We evaluate the errors and compare the result against our goal. Could we decrease the errors by 80 percent with the new tool?*
Now we see that we have reached our goal and decrease the errors by 80 percent. We also saw that the survey did not give us the appropriate results so we had to change the measurement method.



8.2 Conclusion

The spiral model is a methodical approach to analyse the information space or parts of it. It facilitates continuous improvement through the manifold cycle of the spiral. Another important point is that the spiral model can be used on two levels: On the macro-level and on the micro-level. The former is used to analyse the information space as a whole. Questions such as, “what are the most accessed documents?”, “which processes are used and which are not used” etc. can be answered. The latter analyses the information space on a lower level. Here are parts of the information space analysed. Questions such as, “does the search agent work as it should?”, “how do the users use the agent?” or “how do the users navigate through a particular part of the information space?” can be answered.

The spiral model demands us to think about an appropriate measurement method for the analysis of the above posed questions, for example. When an adequate measurement method is identified, then it can be implemented and plugged in to the generic filter pipeline. Latter delivers the results for the evaluation. The spiral model and the generic filter pipeline interact with respect to each other. The former provides the measurement method, the latter plugs it in and delivers results for evaluation.

9. Models, Categories and Properties

Web log files represent user behaviour. Latter is derived from the paths the user passes through the hyperspace. In order to gather the paths of the user, we need to describe the paths as well as the hyperspace. Therefore, this chapter introduces formal models of the hyperspace which are based on graph theory. In introduction, we give a definition and some examples of the hyperspace. Next, we describe the document's attributes. We introduce the document model which defines the relationship between documents. We develop an access model and a security model for the hyperspace. These models describe the access of stakeholders to documents and the security, which the accesses underlie. We define the event model which specifies the dynamic access to documents. Afterwards, the relation between the event model and the document model is discussed. Finally, the website as a representation of a hyperspace is analyzed.

9.1 Utilization models of the Hyperspace

In below, we discuss the utilization models, i.e. views, of the hyperspace. Before we take a look at the views, we give a definition of the hyperspace: *The hyperspace is a sub-space of the information space. It is made up of documents which are linked together*.

- Information medium
- Communication medium
- Collaboration medium
- Marketing medium
- Archive

Example for the usage as an information medium:

At the Institute of Information Technology at the University of Zurich, Dr. R. Riedl has set up the *Zurich's Cynical Theater Guide, a* website which holds information about Zurich's theatre performances. The website has about 10'000 hits per month. The hits are stored in log files, allowing us to analyse the user behavior. Therefore, we are interested in questions such as, "what does the user search for", "how did he get to this site", "how long did he stay", "what did he look at", etc.

Example for the usage as a communication medium:

A big telecom company uses the Intranet as a substitute for its email dispatch. It produces daily reports for the big clients (Bank, Insurance, Airlines). These reports contain data on things, such as the workload of the network, network failures, response time for network failures, etc. They are distributed by a responsible person

on a daily basis to different user groups. The user groups are managers, heads of department, network technicians etc.
In order to automate this process and get rid of the annoying manual routine, the Internet has been involved as follows: The reports were put on the Intranet. The user is allowed to access dedicated reports according to his access rights. The access takes place through the web browser. The responsible person for the reports has to put them only on the Intranet and assign the access rights to each user.

Example for the usage as a collaboration medium:

An example for this is the European project FASME (www.fasme.org). The work group of the different participating countries exchanged their documents through a jointly used, password protected platform.
Each group puts its documentation, reports etc. on the platform. The platform was connected to the Internet and could be reached through a web browser. The reports and charts could be downloaded, processed, and uploaded again to the platform. This permitted decentralised cooperation among the teams without information loss.

Example for the usage as a marketing medium:

Many jobsites such as jobpilot.ch are increasingly used by agencies. Electronic ads are less expensive than newspaper ads, and they can be shown longer to the prospects. They are also available worldwide and the job application can be sent via email.

Example for the usage as an archive:

In this paragraph, we will look at the Internet as an archive from a conventional and a non-conventional point of view. What does the term 'conventional point of view' mean? Documents that have exceeded their validity date are put into the archive. An archive is understood to be a repository for dated documents. Users can search in the archive for old versions. They can search by title, date, version number, etc. We see a practical example from the newspapers that have build such archives. Users can look for old newspapers and read them. However, this service is not free of charge. We see that such archives are read-only, i.e. we can only consume the information that is put into the archive by the owner.
Let us look at the archive, not only as a repository with one owner, but as an open, for all accessible, owner free "knowledge" archive. How would such an archive look like and what would be the advantages and disadvantages?
Such an archive is not "read-only", but "read-write". Each user is allowed to put documents in the archive. Let us take the example of a scientist who writes a paper. In the archive, he finds older documentation and results. Using this information, he continues his research and acquires deeper knowledge which he writes down in the paper. In order that the new knowledge is not lost, the scientist has the opportunity to

put the paper in the archive. In addition, he can set a new link from old reports to this new report that contains new results. What we see here is a link into the future. New document references an old document that it is based on it. The advantage is that a user who reads a paper about a topic of his or her interest can see if there are new papers about that topic or not.
The archive is gradually completed by new research results that are later linked-up. Thus, knowledge flows back to the archive and is available to other users.

In our work, we are going to develop the requirements for the web controlling of information spaces. The space shall support the above mentioned utilisation models with the exception of the archive. First, we introduce an axiom that is of importance:

Axiom

> A generic measure of how well a website supports the implementation of a "project", in the sense of a "dynamic virtual firm".
> The projects must fully meet the customer's needs. If the needs are not satisfied at the end of the project, then this will have negative consequences for all involved parties.

For intranets, obtain:
Project are embedded into the information space and supported by it optimally.

Hyperspace

The content of the hyperspace is made up of documents which are linked together. Furthermore, a dynamic and a static hyperspace exists. Latter is the interlink between documents, formerly consisting in addition documents which are generated dynamically. Figure 32 categorizes the documents.
To obtain a better understanding, we consider for the time being only the static hyperspace. We deal at a later stage with dynamically generated documents with search engines and other tools.
As a further step, the events in the hyperspace will be analyzed and classified. This leads to an event model of the hyperspace.

Figure 32 shows the classification of the hyperspace's documents into following categories:

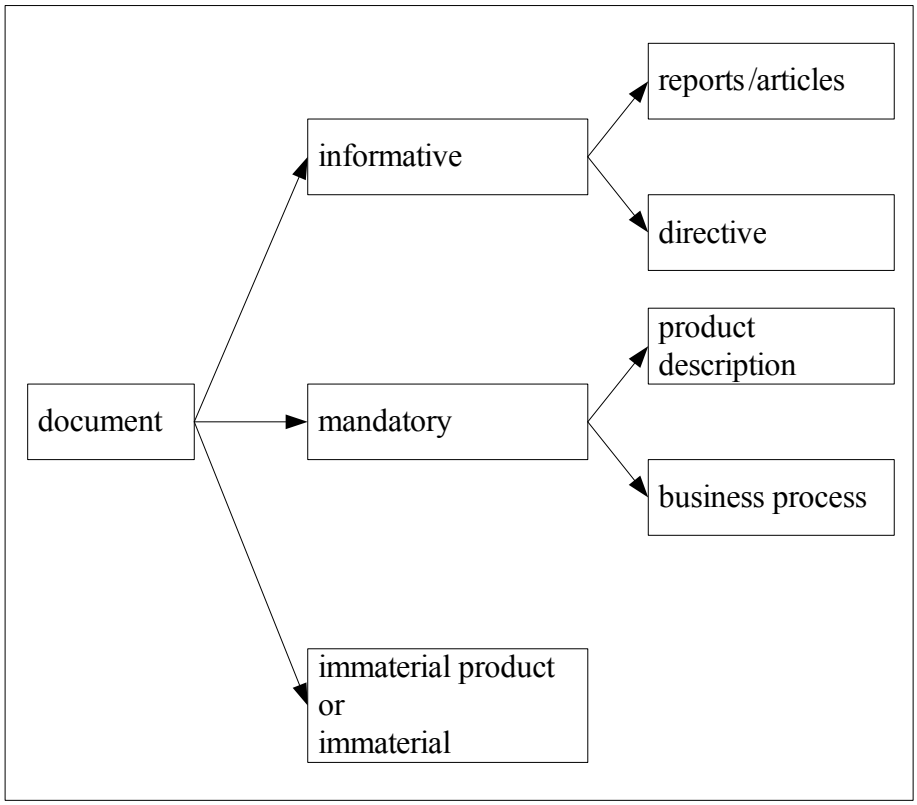


Figure 32 Classification of Hyperspace documents

Reports/articles: Reports about technical, business management etc. areas (e.g. JavaBeans in frameworks, Economy in China etc).

Product description: Information about computers (dell.com), books (amazon.com) etc.

Directives: Mandatory documents which are only on the Intranet and concerned with company issues. Descriptive guidelines like *overtime must be approved by the supervisor.*

Processes: Describe procedures (i.e. provision and purchase of material) in the company.

Immaterial products: Describe documents which the user can use to send messages. E.g. flight booking, arrangement of hotel accommodation, flower ordering etc. Immaterial products form a logical entity and are looked at as a whole.

Attributes

Reports/Articles: *Title*, *Author*, *Date*
Group of interest, keywords (after the abstract)
Dublin core attributes (10) for digital libraries:
Name, Identifier, Version, Registration Authority, Language, Definition, Obligation, Datatype, Maximum Occurrence, Comment.
RDF; XML (i.e. properties of a higher, abtract level, e.g. all documents referenced by document X)
Hyperlink
Role of the author
Status of maintenance

Product description: *Product group, Price, Quality, Manufacturer*
Description
Hyperlink
Status of maintenance

Directive: *Title*, *Author*, *Date*
Group of interest
Validity, Replaces directive
Role of the author
Status of maintenance

Processes: *Title*, *Author*, *Date*
Group of interest
Description business processes, Validity, Replaces process
Role of the author
Status of maintenance

Immaterial Product: *Salutation, Last name, First Name, email, Phone number*
Additional attributes according to the product

Remark:
1. It has to be taken into account in which form the documents are also available. A directive, for instance, can be available as a HTML page or a word document.
2. There are documents which are not in the hyperspace and cannot be accessed through an URL. Nonetheless, they can be send through email and can exert influence to the environment.

Tabular presentation of the document attributes:

	Personal data	Product data	Document data	Additional data
Reports/ Articles			X	X
Product description		X		X
Directives			X	X
Processes			X	X
Immaterial Products	X			

Personal data	Salutation	Last name	First name	Address	Email	Phone number
Immaterial Products	X	X	X	X	X	X

Product data	Product group	Price	Quality	Manufacturer
Product description	X	X	X	X

Document data	Title	Author	Role of the author	Date	Group of interest
Reports/ Articles	X	X	X	X	X
Directives	X	X	X	X	X
Processes	X	X		X	X

Additional Data	Keywords	Hyperlink	Stauts of maintenance	Validity	Replaces previous	Description
Reports/ Articles	X	X	X			
Product description		X	X			X
Directives			X	X	X	
Processes			X		X	X

Table 12 Document attributes

9.2 Document model

We represent the document model as a graph which consists of nodes and edges. The nodes represent the documents and the edges the hyperlinks. An edge connects two nodes and is uni-directional. *A node can have multiple edges leaving it and multiple edges can meet a node*.

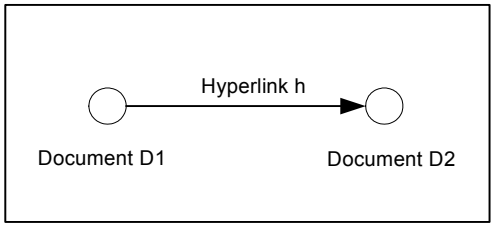


Figure 33 Document model

Let us look at the documents in the hyperspace and their networking. Root stands for the homepage or the portal and is the starting point of the website.

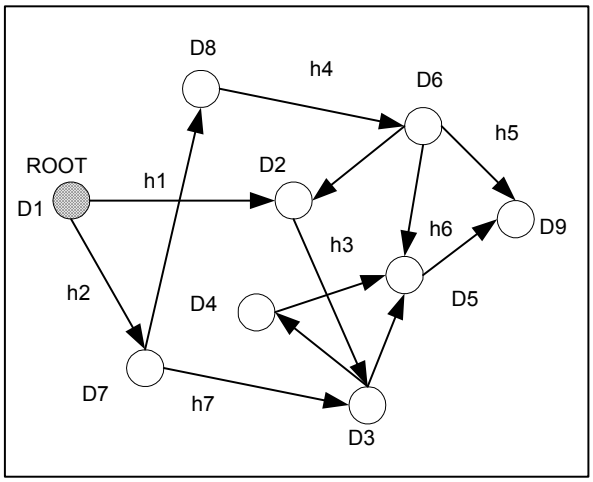


Figure 34 Document network

We conduct an example with the help from figure 34. The graph in figure 34 is not weighted with edge length l=1.

Minimal distance d_{min}(root, D9) = 4

The maximal distance is the longest, intersection free way.

Maximal distance d_{max}(root, D9) = 8

The distance from an arbitrary node D_x to root is:

distance d(D_x ,root) = ∞ , x ∈ {2,…,9}

The distance is infinite if there is no way from D_x to D_y. In our example, there is no way which goes back to the root.

We calculate the distances from the root to the two sub-sites (s. figure 35). The sub-sites are chosen arbitrary.

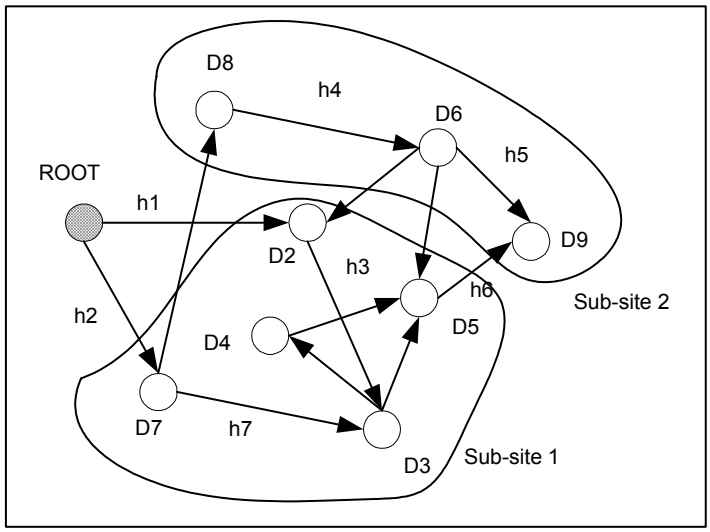


Figure 35 Hyperspace sub-sites

We reduce the sub-sites to nodes (s. figure 36).

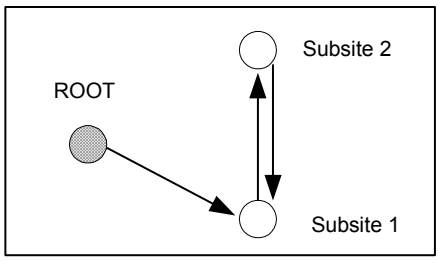


Figure 36 Subsite reduction

distance d(root, Subsite 1) = 1
distance d(root, Subsite 2) = 2
distance d(Subsite 1, root) = ∞
distance d(Subsite 2, root) = ∞
distance d(Subsite 1, Subsite 2) = 1

Conducting a practical example

In order to get practical results, we go through an example with a test group. We ask the test persons to find an information on the WWW. The sought-after information could be, for instance, "*where can I register for the next JavaOne conference?*"

1. We do not provide any information where to find the register page (practical visibility). The test persons has to find their own way.
2. We provide some information where to look at (conditional visibility).

From the results, we extract the time needed for the search, how the search was conducted, and if the person disposed a search strategy and if so which one.

9.3 Access model/ Security model

In the following we look at the relationship between the document and the user. In this context, an access model shall be described and how and which the user is permitted to access documents. We differentiate between active and passive elements. Active elements initiate an action, and we call them the subjects. Passive elements receive an action, and we call them the objects. Therefore, the user as an active element is a subject; the document as a passive element is an object.
Because various subjects have the same rights for different objects, we introduce the term *domain*. We define a domain as a set of access rights. We do not assign the rights to a particular subject, but summarize them into the domain, as described in [Rich85].
For our model, we need two domains, a user domain (subject domain) and a document domain (object domain). The subject domain defines the access rights for the users which reside in it. The object domain defines the access of the user to a document, which resides in it.

Users:
The subject domain must have following attributes:

- access role
- access type
- access group

Access role: We observe in the organizations, the change from a highly hierarchical structure toward a flat, wide hierarchy. We take that into account by introducing roles.
A role describes the function a person has at a given time. Depending on the function, a person can obtain different roles. On a project, he or she can have the role of a project manager (PM), and in other functions he or she can support the sales as a technical pre-sales consultant (sales) or is an IT architect on another.

Access type: Describes the operations which can be performed:

- read: Grants permission only for reading a document
- write: Grants permission for writing to a document. Under writing we understand deleting, inserting and appending of information to the document
- add: Grants permission for the insertion of documents into the website
- delete: Grants permission for the deletion of documents from a website.

Access group: Defines which groups can be accessed. The groups can represent the different departments of a company, like Finance, Sales, Human Resources. A person from the finance department can look at documents of someone from the human resources. However, he or she neither can textually change nor delete that document. However, he or she will only be allowed to add on his on her own initiative documents.

On an Extranet, the groups can represent a company's clients. In this case, the interaction among those would be strictly forbidden.

Document:
The object domain must have following attributes:

- access role
- document group

Access role: Defines which role has access. The set is smaller or equal, but in no case larger than that of the subject domain.

Document group: Specifies which group the document belongs to. According to the above mentioned groups, e.g. Finance, Sales, Human Resources etc.

Remark
It has to be defined by means of the company policy and culture how the interaction among the groups should be handled.

We also have to investigate the relation between users in order to complete the model. This is necessary because we not only have "normal" users but also "superusers", who are the administrators.
A superuser (e.g. administrator) must have following rights:

Propagation of rights: A superuser can hand his rights to a "normal" user. Thus, the user becomes a superuser.

Read (of websites): The superuser can disable parts or the whole website for particular users.

Assignment of operations (read, write, add, delete):
The superuser grants access rights to a user for a document.

We also introduce the principle of the "owner" and the "user". The user is the one who uses digital entities (document, website). The owner is the one who owns digital entities and can also be a user at the same time.
The owner can grant or deny access rights for his documents to a user. The superuser is owner of all entities and stands on the top level of the hierarchy. Thus, he or she has ownership rights to all entities.

We see following scenario in figure 37:
The user who resides on the subject domain with following attributes, access role is "PM", access type is "delete", and access group is "IT", deletes a document from a website. He can do so because the document are the appropriate attributes. The access role is "PM" and document group is "IT".

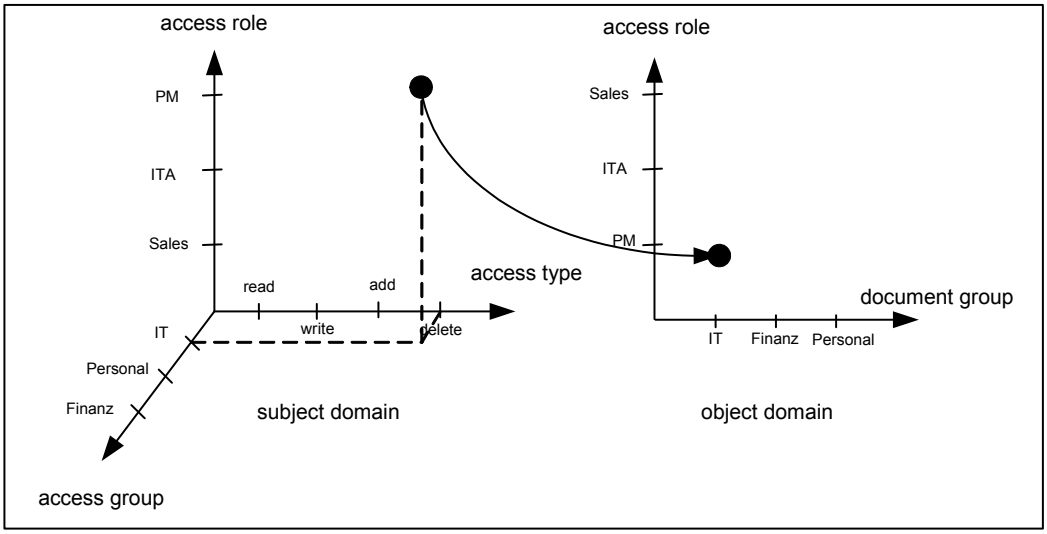


Figure 37 Access/ Security model

In figure 38, an IT architect unsuccessfully tries to insert a document into a website. The reason is that the object domain only allows the access from a project manager (PM). The access role of the object domain is set to "PM".

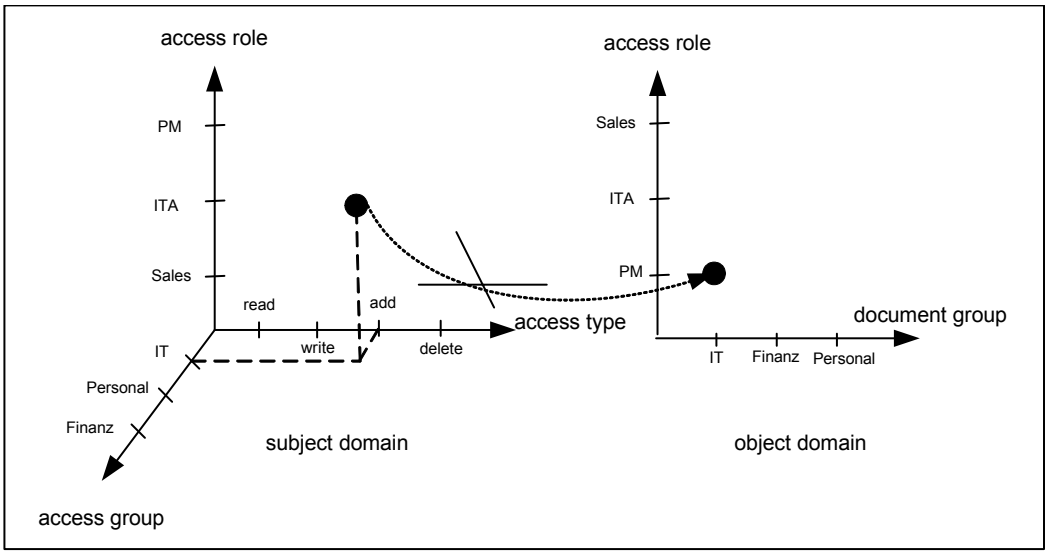


Figure 38 Access failure in the access/ security model

We also have to consider the rights allocation from an organisational point of view. Since the capability model resorts to the lower layer, we have to ask ourselves which guidelines the top management has to formulate in order to allow a correct implementation. The guidelines contain the rights for the classes[99] and roles[100]. In addition to the rights, a role also holds duties which are stated in the guidelines as well.

[99] Indicate to above mentioned classification like directives, processes etc.
[100] Roles are realized by persons. A person can take up different roles.

Thus, the management formulates on a higher and at an abstract level the access rights. These apply to classes and roles and define the blueprint for the downward breaking to the capability model. Directives, for example, are instances of the class.

A further job of the management is to provide the provision of the procedure for changes of the guidelines, since changes of the guidelines also entail changes of the rights between classes and roles. *Are the rights only allowed to be changed when primarily the according guideline resp. guidelines had been changed*, then we entail any flexibility, but gain overview and control.
Let us explore the case where the rights between the classes and roles have been changed without changing the guidelines. Such a proceeding allows a particular flexibility for ad hoc solutions. Until the solution finding and implementation, people can work fast and unbureaucratic with each other. If the work is finished, then the rights are suspended. Such a flexibility can also be misused by defining the emerging spontaneous processes such as ad hoc. The guidelines are undermined and consequently invalid.

Reproducibility of information

In the hyperspace, the reproducibility of information is not guaranteed. In order to guarantee it, a Version Control Management System (VCMS) must be utilised. Each document is versionised before changed. A problem, which can be seen by now, is the huge data which accumulates over time.

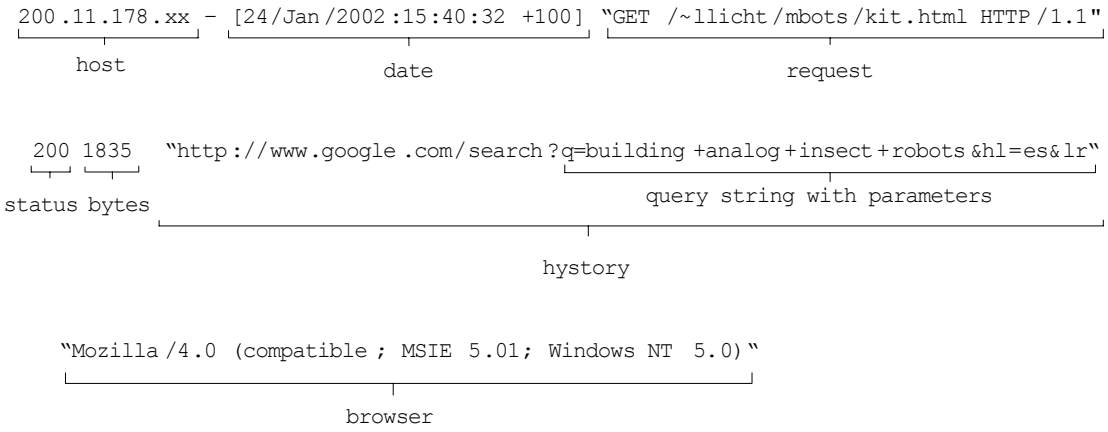
9.4 Event model

For the event model, we survey the requests on the web log files. The requests look as follows:

host ident authuser date request status bytes

host: IP address or domain name of the user
ident: Identification information of the user
a*uthuser*: User ID, if the request is for password protected document
date: Date and Time of the request, format:
[day/month/year:hour:minute:second tzoffset]
request: Request of the user
status: Status code which is sent back to the user
bytes: Amount of bytes which were sent back to the user (without headers)
history: Indicates the previous Webpage (optional)
browser: Type of the used web browser, eventually with the operation system

Example of a request



We regard all Internet protocols in the application layer (HTTP, FTP, Telnet, SMTP) and derive a generic event model for the Internet communication. More concretely, we take a look at the HTTP protocol and at all events at the FTP protocol.

The event model consist of following attributes:

1. Source
2. Time
3. Target

Each of these attributes has further attributes:

1. Source:
   - IP Address
   - Session
   - Authuser
   - Ident
   - Access context
2. Time:
   - Date
   - Time
3. Target:
   - protocol://host:[port]/resource-path (URL[101] for HTTP protocol)
   - access method (e.g. GET for the HTTP Protocol)
   - Properties of the request object

Remark: The attribut *session* is an artificial one since it is calculated from the attributes IP and Time.

The model is based on the client/server principle.
The client is an entity which initiates an action. Its attributes are time[102], IP address, and all other relevant information which can be derived from a request.

[101] Assumption: a logfile for a domain name
[102] Server time approximated for the client site.

The server is an entity which receives the action. Its attributes describe the context of the actual text.

"Push" Konzept

A ticker or a banner advertisement, as we see, is updated periodically. The update comes from an applet or a script in the web browser. After a certain time, a request is sent to the webserver which sends back the updated values. From a technical point of view, a GET, resp. a POST method is evoked. What looks like a push is a periodically evoked GET/POST method from the browser.

Frames

Another property in which we have to consider is frames. In a frame, the user has the possibility to navigate in the HS without leaving the main web page.

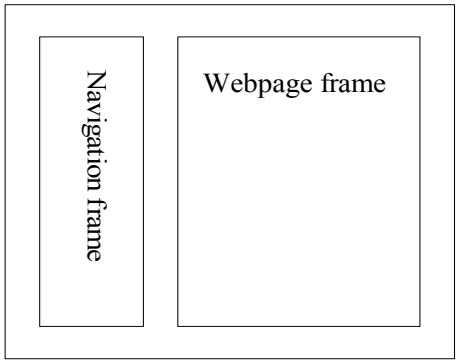


Figure 39 Frame layout

This is possible because there is a navigation frame on the left hand side. The navigation frame has different hyperlinks. A on a link shows the appropriate view of the hyperspace on the right hand side.
Thus, we have within a super event sub events. The access to a webpage represents the super event, and the navigation in the webpage frame represents the sub events. Hence, we also derive super and sub sessions.

Protocol dependencies

In figure 40 we show the network protocols. Our interest lies in the application layer and concerns the HTTP protocol. They coexist and do not influence each other.

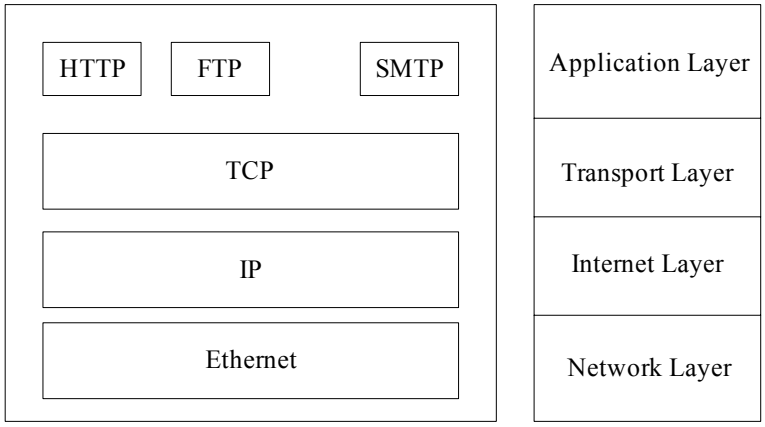


Figure 40 Source [Tane02], Network protocols

Representation of the event model

We represent the event model as a graph which consists of two nodes and one edge. The two nodes represent source S and target T, and the edge represent the event e with time t. The edge connects the source with the target.

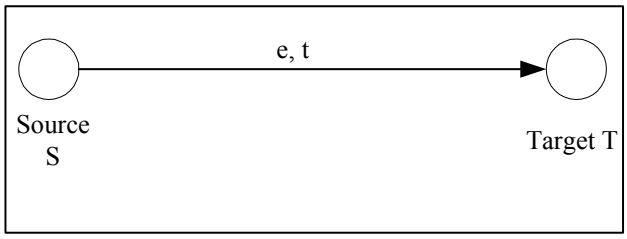


Figure 41 Event model

Let us look at a session which looks like following:

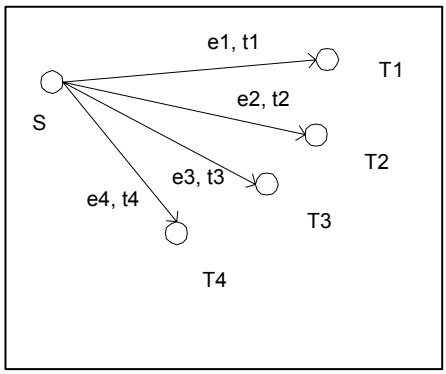


Figure 42 Session model

In a session, the source accesses several targets. Each access, event e_n, is invoked at a certain time t_n.

Source $S \rightarrow \{T_1, T_2, \ldots, T_n\}$
Session $s_{[t1,tn]\,[e1,en]} \rightarrow \{S, T_1, T_2, \ldots, T_n\}$

We need the definition of session because the web log files only give the requests to the server. With the above definition we artificially create sessions for different sources. Latter is in reality users which access web servers. The Generic Filter Pipeline chapter explains in detail how sessions are created.
The filter pipeline implements the method *createSession,* which joins the requests into sessions.

Path in the document model

In the document model, a user would describe the following path. We take the document representation of figure 34, for instance, and put the path which the user describes above in the document model.

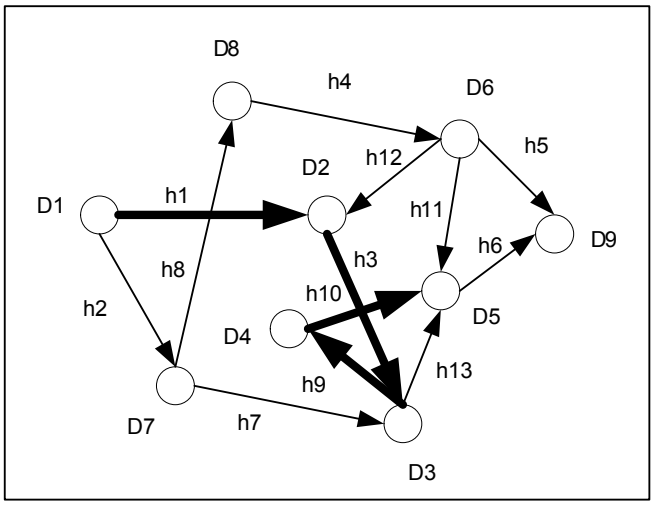


Figure 43 Path in the document model

The path is: Path $P \rightarrow \{D_1, h_1, D_2, h_3, D_3, h_9, D_4, h_{10}, D_5\}$. It consists of the
Nodes $N \rightarrow \{D_1, D_2, D_3, D_4, D_5\}$ and the
Edges $E \rightarrow \{h_1, h_3, h_9, h_{10}\}$

9.5 Relation Event model – Document model

The event model describes a path on the document set, thus, most of the paths lie on the link path of the document model. If the different paths are put on upon the other, then we can read out which sub-paths are most frequently used. Out of it, we derive

different patterns. The patterns characterize the profile of a certain user group. Therefore, we identify the profile of a particular user as well as that of a user group.
→ Personalisation, classification (clustering)
→ User behaviour
→ Colouring of the path within the document model
→ Re-engineering of the Intranet

By the use of colouring, we create a quotient space.
We introduce two kinds of colouring for graphs. For the textual kind of properties, we use attributes for the user behaviour numbers. The user behaviour can represent the "n-th" webpage access of a session, or "the number of users which visit resp. leave the webpage", or the "n-th request word of a user", etc.

In figure 44, different attributes of nodes are displayed through colours.

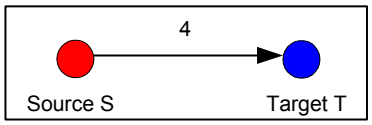


Figure 44 Quotient space

Figure 45 represents the coloured graph of figure 34. The colouring is chosen arbitrary. Nodes with the same attributes have the same colour.

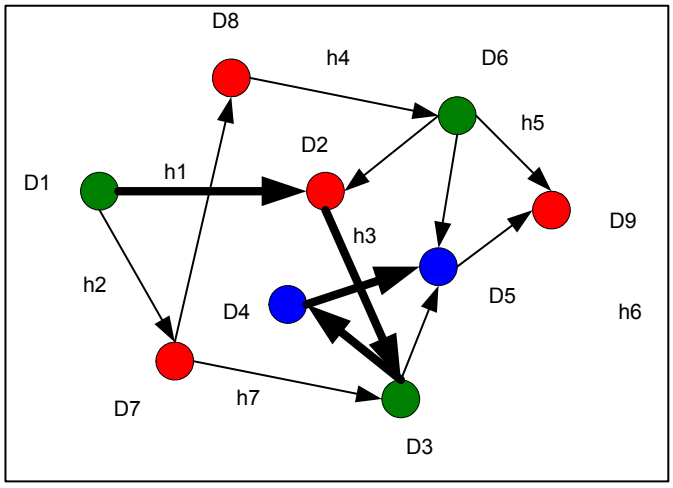


Figure 45 Path in the quotient space

Figure 46 shows the collapsed Graph of figure 45 to the three in their properties different nodes. The nodes in figure 46 now represent the classes of the site. The numbers on the edges represent the order of the visited nodes.

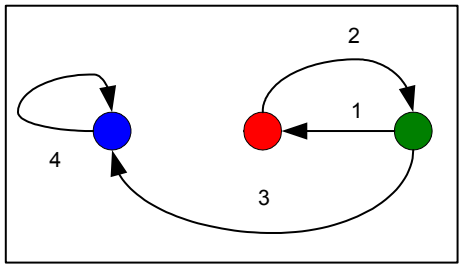


Figure 46 Collapsed graph

Let us look closer to the edges. An edge represents a hyperlink. The two ends are the source and the target (s. figure 33).

1. At the lowest level, the start point is the link itself. The end point is either an anchor or a webpage.

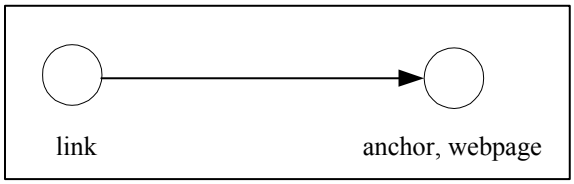


Figure 47 Link-anchor relationship

Implementation of the link in HTML:
Anchor: *<A href="* *www.company.com/whitepaper.html#chapter1* *">*
Webpage: *<A href="* *www.company.com/blueprint.html* *">*

2. If we move up an abstract level, we no longer look at the link itself, but also at its surrounding context. This means that the source represents a webpage. The target is either a webpage or a website.

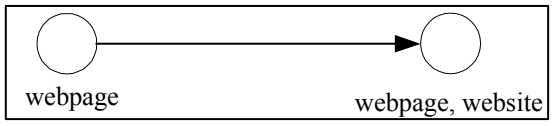


Figure 48 Webpage-webpage/website relationship

3. If we move up another abstract level, then the nodes, source and target, represent both websites from one website point to another website.

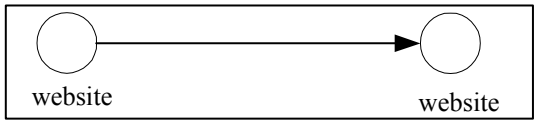


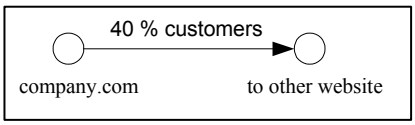
Figure 49 Website-website relationship

The three abstract level give us the chance to analyse user behaviour on different levels. On the highest level (s. figure 49), we see if the user has left the website add

the amount of the users who have visited the website. The amount is apparent from the weighting of the edge.

An example:

> A company sells a product online and realized that despite the huge attendance, only a few users buy the product online. An explanation could be: "*The user is not interested in the product.*"
> If we analyse the graph, we see that a lot of visitors leave the site.



> On a more precise consideration of the website we see here is that there is a product description before the online order link which points to *amazon.com*. We conclude that the users go to amanzon.com before buying in order to see an additional description of the product. At amazon.com, they get distracted and do not end up returning again. Consequently, these are lost customers.
> The company must remove the link so that the visitors go directly from the product description to the online order without distraction.

Through the analysis of the graph, we can draw such conclusions. It has to be considered that such conclusions are not universally valid, but must be put in the context of a particular situation.

9.6 Website characteristics

A website is a set of documents which are logically connected to each other and accessible through the same homepage (www.ifi.unizh.ch). If the user goes to another website (e.g. www.unisg.ch) then we say that he or she has left the website. The webpages of a website are stored in directories. The file system is built up hierarchically and is described in a tree structure. At the top, is the homepage. Starting from it, the user can navigate deeper into the website, whereas the next lower levels always represent logical entities.
For an example, we take the website of the Institute for Information Technology[103] (IFI) at the University of Zurich. The homepage contains links which give the users a fast overview of what resides behind them. Some examples are, a link to the staff, to the library, to the research groups, etc.
If we lay down the tree structure, we get figure 50. The structure shows in which directories the documents stand.

[103] Former name of the department. The web structure refers to the old design.

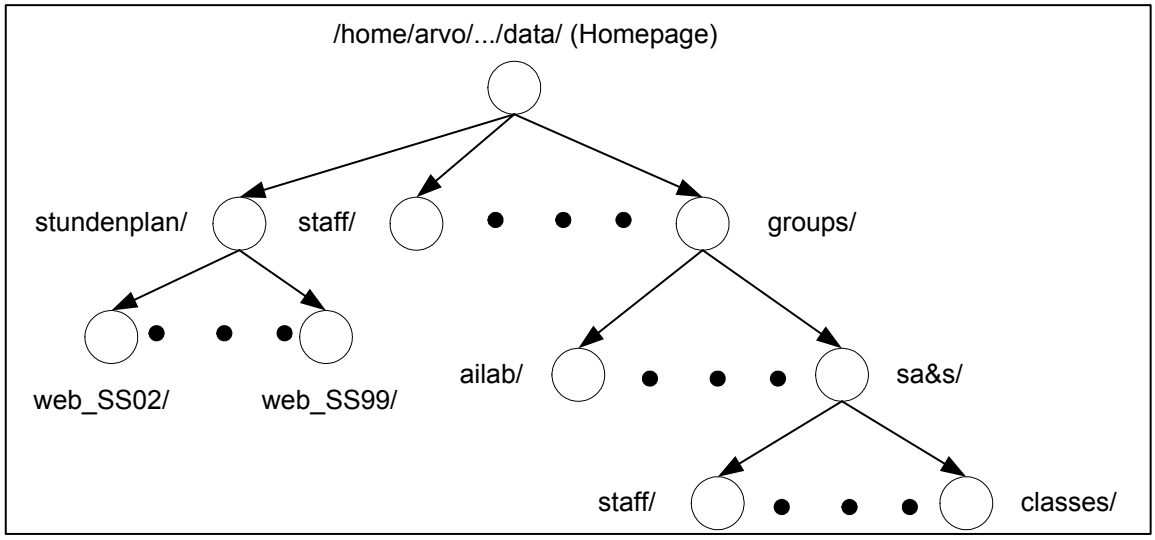


Figure 50 Tree structure of a website

If we lay down the link structure, we see that it includes the tree structure. We also observe that links leave the tree structure and point to pages in other directories.

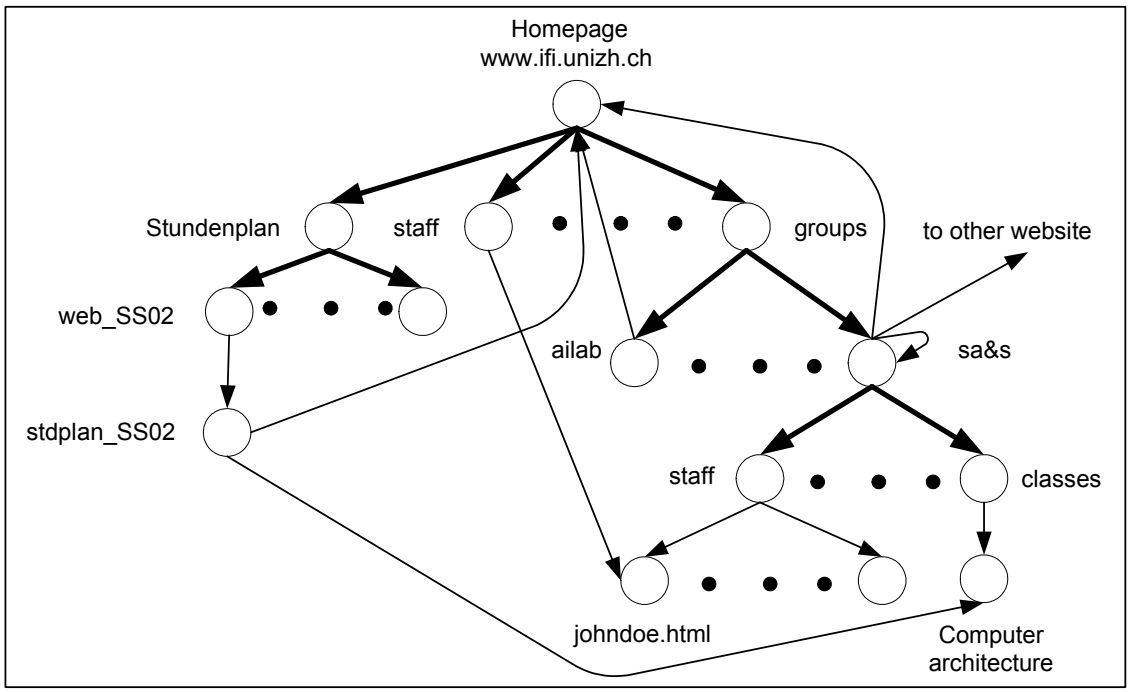


Figure 51 Link structure of a website

For the user, the link structure matters. Apart from the homepage, he will not reference a document by entering the URL manually. He concentrates on the hyperlinks within the document. We see in figure 51 that links point to their own document. Technically, this happens by defining one or more anchors in the

document. This allows a fragmentation of the document in different sections. If one clicks on a link then the page goes to the appropriate section.
Example 1:

> *<A href="#introduction">Introduction</A>*
> *<A href="#chapter1">Chapter 1: Sockets</A>*

The user jumps to the introduction or the first chapter within the same document. One can also define links which point directly to a section on another document.
Example 2:

> *<A href=" www.company.com/whitepaper.html#chapter1">Chapter 1: Sockets</A>*

The user jumps directly to the first chapter of the document, "white paper", which can be found on the website *www.company.com*. He will not get the chance to see the introduction.

What are the consequences in this regard? A *tunnel view* which can have annoying impacts under certain circumstances is generated. We introduce an example for illustration.

> If an employee jumps directly to a section of a directive, he will not see what is above that section. Since it is a directive and the employee must take it seriously, he has to be informed correctly. If the employee reads only this section without being concerned about the part above, he has a good chance to miss important information. It could be stated that: "*This directive is no longer valid*" or "*This directive is replaced by…*"

Conclusion: The view of the user is distracted in such a way that wrong information is seen as right.

Remark
We have three structures on the set of the documents, which therefore, become vectored graphs. There are the hierarchy of the file systems, the link graph, and the path graph.

Classification of websites

Figure 52 shows our classification of websites.

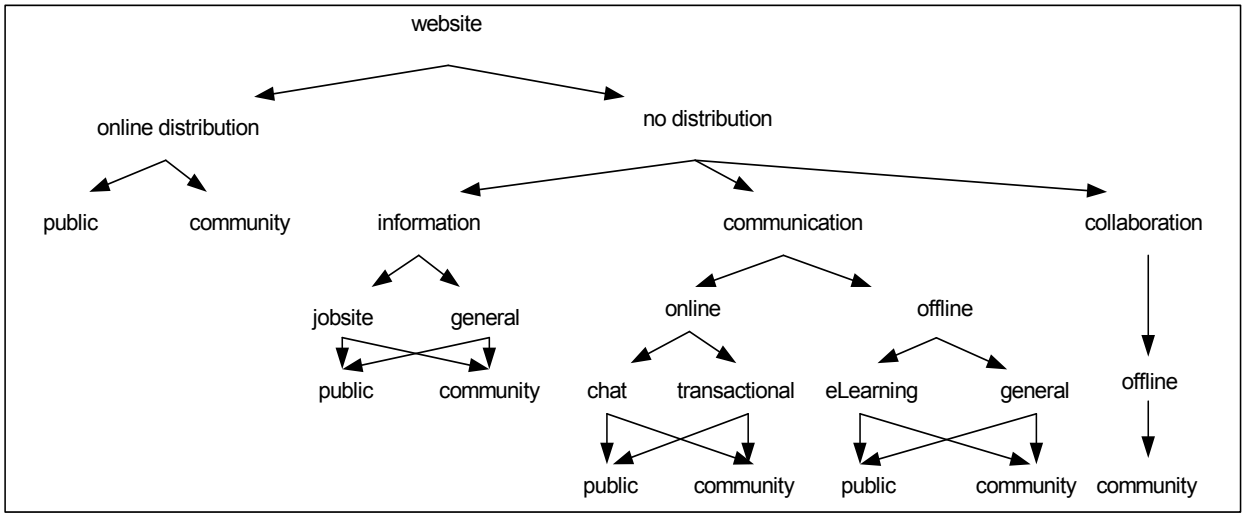


Figure 52 *Website classification*

Legend

Distribution: Sales, marketing over the WWW. A classical example for public distribution is amazon.com. An example for internal distribution is e-Procurement within a company.

Information: We basically distinguish between two main lines. Websites, which put information mainly about job offers on the WWW. Such jobsites can be either public (topjobs.ch, jobpilot.ch), or only within the company available.

Communication: Under online communication fall the popular chat rooms. A particular attention must be given to the company's internal chat rooms. It is not about chat among co-workers, but about professional teamwork. Chat rooms are popular at start-ups in order to cut costs for phone, fax, and etc.
Transactional site can be found at financial portals or banks.
Training over the web increasingly gains importance. eLearning promises flexibility and cost reduction.

Collaboration: In a collaboration, different parties work together to achieve a goal. The parties can work geographically and temporally independent from each other.

In our work, we give special attention to the following sites:

1. UBS: Intranet site
2. FASME: Project site
3. Luchsinger Mathematics: Jobsite
4. Institute of Information Technology: Website (public)
5. Zurich's Cynical Theater Guide: Website (public)

We classify the above sites on the basis of our classification schema (s. figure 52).

1. UBS: no distribution → communication → offline → general → internal
   → The UBS Intranet is a website for internal communication
2. FASME: no distribution → collaboration → offline → internal
   → The FASME-site is a platform for the collaboration of the project co-workers
3. Luchsinger-Maths: no distribution → information → jobsite → public
   → Traditional jobsite with additional information
4. IFI: no distribution → information → general → public
   → Public website with information about university studies and links to other institutes
5. ZZT: no distribution → information → general → public
   → Public website with information about theatre performances

Contemporary websites have emerged complex information spaces. These information spaces can no longer be classified into the one or the other category. In fact, they are a network of collaborating sub-sites.
If we survey the website of a big universal bank, we see (s. figure 53)

- The site has general information about the bank (sub-site 1)
- The site is eBanking enabled (sub-site 2)
- The user can look at job offers and apply online (sub-site 3)

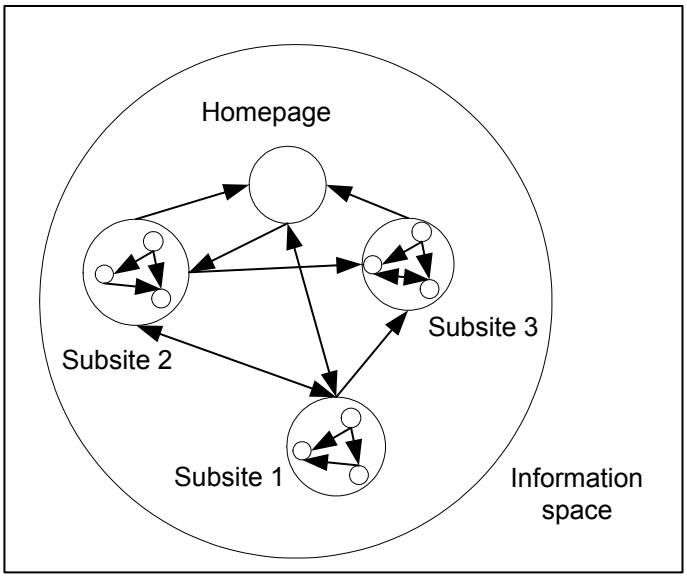


Figure 53 Information space

User groups

These kind of users can be derived on the basis of the log files. Are they internal or external users? Do they look for something in particular (e.g. product -, company description)?

9.7 Conclusion

We have developed formal models which describe the hyperspace and the static and dynamic accesses in it. Based on these models, a better understanding of stakeholders' navigation is acquired. In addition, models support the definition of new evaluation methods for the analysis of user behaviour. These methods can be implemented and plugged in to the generic filter pipeline.

10. Generic Filter Pipeline

In the previous chapter we introduced a model framework to represent user behaviour. In this chapter we introduce an analysis tool which computerizes the user behaviour from the model framework. We name this tool the generic filter pipeline. It enables to do arbitrary evaluation of access event streams of variable filter stacks. The generic filter pipeline is an open architecture. Different modules can be plugged in it to enhance the evaluation functions. The reasons why we need an open architecture is twofold: First, the spiral model facilitates new evaluation methods, and second, the formal models enable the implementations of new functionalities. The evaluation methods can be transformed into new functions which are based on the formal methods. Therefore, the pipeline is a dynamic evaluation tool. In the following sections, we describe the method to our pattern search. First, we identify the data which we are looking at. Second, we explain the tool which uses the data for further manipulation and outcome. Third, we give a conceptual realisation of our method and lastly, we end with a practical realisation. We also give a short explanation for clustering.

10.1 Introduction to the Generic Filter Pipeline

To start, we look at the data we are going to analyse. The data we look at consist of requests from the web server log files. The requests have the form as described by the *Event model*[104]. We also group requests into sessions. A session is a set of requests from the same origin and with a maximum time difference of Δt (Δt is a variable and therefore can be set by the researcher) from $request_n$ to $request_{n+1}$
($1 \leq n \leq N$, N: set of requests).
To embrace the term clustering, we understand the method of grouping. Clustering groups similar objects into classes. The classes are disjoint and the union is equal to all requests or sessions, respectively. In other words, the classes form a partition. In our explanation, the objects are requests and sessions.

10.2 Pattern search

In this paragraph, we describe the way we look for patterns in our log files. We pursue two approaches. A general one and a particular one.
First, we explain the general approach. In this, we define a general schema which applies to all websites we look at. This schema must be unambiguous for the set of websites we analyse, cum grano salis. To describe such schema, we have to find criteria which are valid for the set of websites. Hence, such classifications can be based on, just to name a few,

[104] Chapter 9.4, p. 98.

a) website indexing based on content
b) link structure and —depth
c) category groups

For a better understanding, we assert an example. Let us assume that we have two websites A and B. We take the criteria b to compare both websites against that criteria. The questions we can ask are: How are the links organized on a particular page? Do the websites have a flat or a deep hierarchy?
The drawback we face with the general schema is that not all criteria have the same relevance to each website. Some can be more relevant than others to a particular site.
Second, we explain a particular approach. The principal purpose is to find typical structures within a website. The criteria for the structures can be:

a) characteristic patterns
b) pure site structure

We apply this criteria individually to each site. The problems we first face are, whether we can find such structures. The second is how relevant the comparison of those structures is against each other.

Steps of the pattern search process

We implement a module which parses the logfiles and fills the tables in our database.

Step 1: We generate sessions from the database entries. It is necessary to do this because the logfiles just give us the requests from the users. To define sessions, we choose an arbitrary Δt. On the one hand, the size of Δt is chosen by experience of the researcher and on the other hand, also a reasonable upper time limit of the data set is determined. If the requests are within the range of Δt then they belong to the same session, otherwise a new session starts. After having set Δt, it remains fix for the succeeding analysis.[105]

Step 2: At this stage, we create the session file F. The session file consists of sessions with its requested accessible objects (rao).

Set Requested Accessible Objects
RAO:= { rao | rao $\in$ logfile, $1 \leq$ rao $\leq N$ },
rao: requested accessible object, N: natural number
Session file $F = (f_n)_{1 \leq n \leq N}$
E.g.: Session file F = {0, 1, 2, 7, 1, 0, 3, 1, 1, 0, ……,0}
session 1 session 2

[105] For the definition of session see chapter 9.4, p. 101 et seq.

The sessions in the session file are delimited by zeros. The zeros in the file mean that the user is coming from or leaving to another site. In addition, the file starts and ends with a zero.
The above entry shows session 1 and session 2 with its rao's 1, 2, 7, 1 and 3, 1, 1 resp.

Step 3: After creating the session file, we look for F^{next}.
The file F^{next} contains the requested accessible objects which are next to the requested accessible object n. We create the F^{next}_n ,"request object next file", for each request object.

Request next file $F^{next} = (F^{next}_n)_{\ 1 \le n \le N}$
E.g. : We conduct an example where we illustrate the creation of the request next file F^{next}_n. In our example, we create the F^{next}_1 file. Hence, the requested accessible object n is 1. We start with the above introduced session file F.
Session file F = { 0, 1, 2, 7, 1, 0, 3, 1, 1, 0, ……,0}
In the file, the rao 1 is underlined. The rao which are next to 1 are gray highlighted and it forms the request next file F^{next}_1. Thus, to create the F^{next}_1 file, we take the rao's which are next to 1, thus, the highlighted ones.
Request next file F^{next}_1 = {2, 0, 1, 0}, rao n = 1.

Step 4: From the F^{next} session file, we create the transition matrix (Markov matrix)[106].

The Markov matrix represents the transition probability from request object i to request object j. The probability is a_{ij}.
Markov matrix $M = (a_{ij})$, $i, j \in RAO$.

	a_0	a_1	a_2	…	…	a_{n-1}	a_n
a_0	a_{00}	a_{01}	a_{02}	…	…	a_{0n-1}	a_{0n}
a_1	a_{10}	a_{11}	a_{12}	…	…		
a_2	a_{20}	a_{21}	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
a_{n-1}	a_{n-10}	a_{n-11}	…	…	…	a_{n-1n-1}	a_{n-1n}
a_n	a_{n0}	…	…	…	…	a_{nn-1}	a_{nn}

Table 13 Markov Matrix M

a_{00} is always zero since there are never two or more consecutive zeros.

We conduct an example to demonstrate the creation of the matrix. The example is based on the session file F (see step 2).

[106] See [Hast89].

Starting from the request next file F^{next} , we will calculate the a_{ij} elements of the matrix. We have defined above $F^{next} = (F^{next}_n)_{1 \leq n \leq N}$. We choose n = 1, which is equivalent to requested accessible object = 1. Thus, we calculate the probabilities from the rao 1 to the other rao's. Since we have taken rao 1, we must take F^{next}_1.
With $F^{next}_1 = \{2, 0, 1, 0\}$ (s. above), we create the matrix elements a_{10} a_{11} a_{12} and a_{13}.
Let be $f^{next}_{ij} := |\{j \in F^{next}_i\}|$.
The matrix element $a_{ij} := f^{next}_{ij} / |F^{next}_i|$

E.g. for $F^{next}_1 = \{2, 0, 1, 0\}$
$f^{next}_{10} := |\ 0 \in F^{next}_1\}| = 2$
$f^{next}_{11} := |\ 1 \in F^{next}_1\}| = 1$
$f^{next}_{12} := |\ 2 \in F^{next}_1\}| = 1$
$f^{next}_{13} := |\ 3 \in F^{next}_1\}| = 0$
Hence,
$a_{10} := f^{next}_{10} / |F^{next}_1| = 2/4$
$a_{11} := f^{next}_{11} / |F^{next}_1| = ¼$
$a_{12} := f^{next}_{12} / |F^{next}_1| = ¼$
$a_{13} := f^{next}_{13} / |F^{next}_1| = 0$

The Markov matrix M would be

	a_0	a_1	a_2	a_3
a_0	…		…	…
a_1	0.5	0.25	0.25	0.0
a_2	…	…	…	…
a_3	…	…	…	…

E.g. A real example from our considered website. Markov matrix

M_{real} =

	rao_0	rao_1	rao_2	…	…	rao_{n-1}	rao_n
rao_0	0.0	$1.672 \cdot 10^{-3}$	$1.672 \cdot 10^{-3}$	…	…	$3.344 \cdot 10^{-3}$	0.0
rao_1	1.0	0.0	0.0	…	…	0.0	0.0
rao_2	1.0	0.0	…	…	…	…	…
…	$5.609 \cdot 10^{-1}$	…	…	…	…	…	…
…	…	…	…	…	…	…	…
rao_{n-1}	...	…	…	…	…	…	...
rao_n	0.5	…	…	…	…	…	…

Table 14 Example of a Markov Matrix M

E.g.: The transition probability a_{01} is $1.672 \cdot 10^{-3}$% (in the matrix above gray highlighted). This means that the probability that a user goes from request object 0 to request object 1 is approx. 0.17%. The results are based on the set of data we took into account.

Step 5: We calculate the Eigenvector V of the Markov matrix.

The Eigenvector represents the balanced state of the system. It should be similar to the relative occurrence of the request objects. Only then it makes sense to look at the website as a Markov-like system.
In addition, the Eigenvector V is an abstract measure for the importance of the webpages.
Eigenvector V = (v_1, v_2, …., v_n), n = number of request objects, to the Eigenvalue x = 1.
We calculate the Eigenvector V with the routine described below. The Vector x and Matrix M are normalized.

M: Matrix
x: Vector
loop:
$x \bullet M =: x_{new}$
$x := x_{new}$

After N iterations, the system moves into a balanced state. A balanced state is characterised by the fact that the difference between the vectors x and x_{new} is 0. The vector x becomes the Eigenvector of the matrix M.

E.g. Eigenvector V = {
$3.790 \cdot 10^{-1}$
$6.401 \cdot 10^{-4}$
$6.401 \cdot 10^{-4}$
$6.401 \cdot 10^{-4}$
$6.401 \cdot 10^{-4}$
$1.735 \cdot 10^{-1}$
$4.481 \cdot 10^{-3}$
…
…
v_{n-1}
v_n }

The Eigenvector represents the probability of the duration of stay. The sum of its elements is equal to 1.

10.3 Description of the pattern search process

In this section, we describe the pattern search process. In our method, we emanate from the basic events which occur in our system. These are the requests.[107] From

[107] See Event model, chapter 9.4, p. 98.

those requests, we append context information and describe their relationship to each other. We also need to know the entire event stream as well as what occurs shortly before and shortly after a request.
We name the process of adding context information to the request, *attribution*.
The attribution is generated by the particular request and the context information.
First, we add attributes from the request itself to the context. The main step of the context classification is the definition of a partial order of events, which in that is compatible to a temporal order, so that

if $a > b$ (a, b are requests),
$\Rightarrow$ timestamp(a) > timestamp(b)

holds. And this partial order has the special characteristic that defines an equivalence relation with following characteristics:

$a \Leftrightarrow b$ iff
$a < b$ or $b > a$

is defined on all elements in the partition which are, in turn, defined on the equivalence relation.
If the partial order is a local linear order, then we call the individual partition, a session. Then we demand that the appropriate requests of a session have the same IP address.
The local context is defined through the partial order. The global context is defined through statistics (s. Zipf function in [Ried99a]).
The attribution in which we undertake has three sorts of attributes which are ones we generate for the requests themselves, from the local context and from the global context. The consequences are:

1. Generation of large load. Partition (session, IP) of all request altogether.
2. Results in pairwise relationships between requested accessible objects.

We implemented a tool with which we analyse the events of the system. The events are requests which were made to the system. For our research, we consider a set R of requests.

The major task of the tool is the attribution of the data which flows through the system. Former is component oriented and each level is independent from its predecessor. The components can be put in a row and hold input parameters. These parameters define the attribution of the data and they award the components with flexibility.

As mentioned above, requests flow into the tool. At the first level, they are parsed by IP address, time, request object etc. and are written into a database. The second level creates first statistics. Level three enables manual pre-clustering of the requests. This clustering is a colouring of the requests, which is latter being assigned an attribute. Level four is a filter which generates sessions from the requests. Hence, we introduce a partitioning of the requests. Filter five, analog to filter one, parses the sessions and reads the attributes IP address, begin of session, end of session, length of session, requests objects. Level six creates further statistics (e.g. amount of requests) based on the sessions. Filter seven, analog to filter three, enables pre-clustering of the sessions.

The complex filter type 1, level eight, generates different views of the requests and the sessions based on the colouring which has been made so far. The last level, level nine, allows more complex queries of the data.

Graphical illustration of the pattern search process

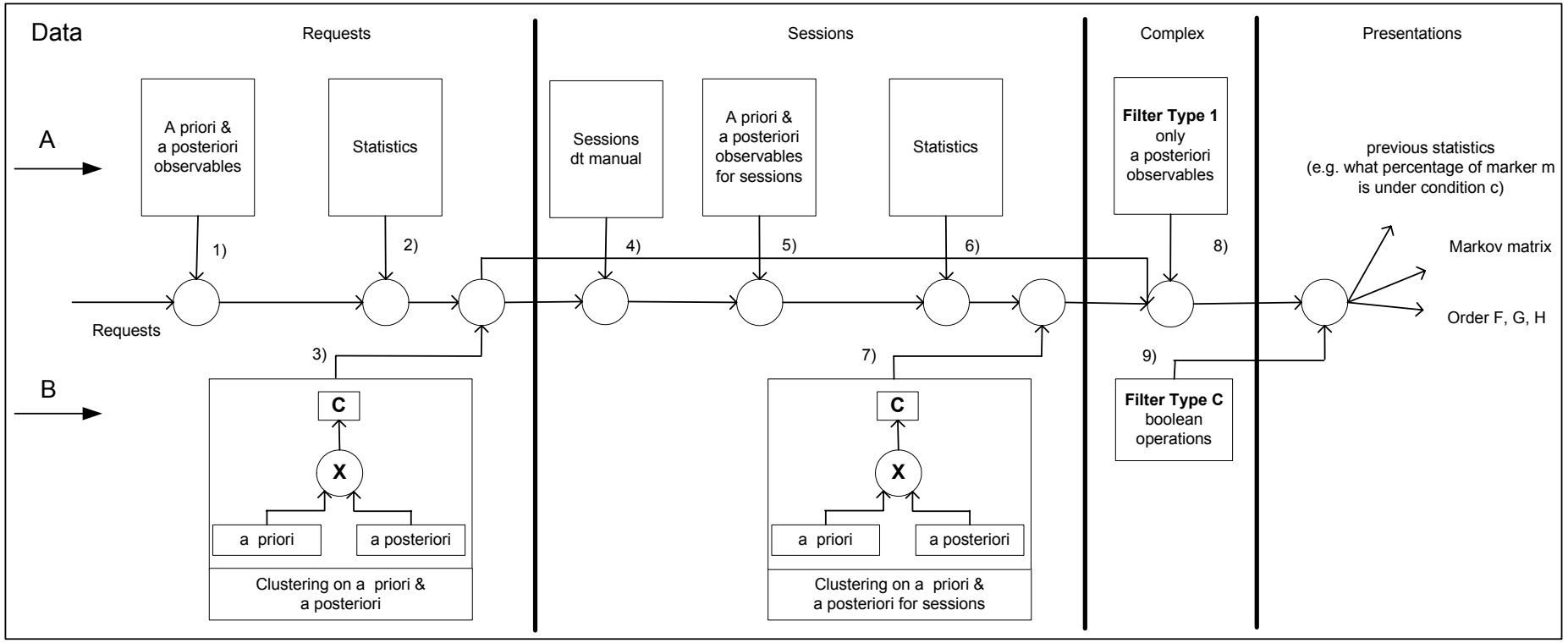


Figure 54 Generic Filter Pipeline

Legend

A. All the steps are automated deterministic processes. Tools support us in manipulation and calculation of data.
B. All the steps are random, manual processes. We interact manually with the system.

Preprocessing Task

The preprocessing tasks, we need to perform before we are able to analyze the data, are the following:[108]

1. The web log file must be cleaned from crawlers, jpeg, gif, pdf and ps files. This is done with the method: *filterLogfile*
2. Sessions must be created from the requests. This is done by first identifying all requests which belong to a user and by having an interval equal or less to Δt among each other. This is completed with the method: *createSessions*
3. After that, attributes and parameters can be identified.

The particular steps of the pipeline are documented below. Each number corresponds to the specific number in the pipeline.

Attributes and Parameters

1. IP, Time, 1 Request object, HTTP Request, History.
2. n-th frequent page, $\in A_i$ (page A which is referenced I times).
3. manual clustering C_i on *a priori* and *a posteriori* observables for requests.
4. Session file.
5. IP, BOS, EOS, LOS, Request objects (IP, BOS are a priori; EOS, LOS are a posteriori).
6. Number of requests, n-th length session.
7. manual clustering C_i on a priori and a posteriori observables for sessions.
8. Information which is only collected from a row in the session file.

8A. with a priori and a posteriori observables extended session file.

9. all Boolean operations.

In the following, we give the definition of queries for the pipeline which is given in EBNF[109]. This queries activates the analysis of the web logfiles which are pushed through the pipeline. The input of the pipeline is the requests on the web logfiles.

Definition of Queries

StartEvaluation → method, parameterList
parameterList → {} | parameter, parameterList | request_object_set
parameter → value | value, parameter | parameterInterval
parameterInterval → date, date | date, date, parameterInterval
method → {t_m | t_m is
getBeginAndEndOfSession,
getFirstRequestObjectInSession,
getLastRequestObjectInSession,

[108] See chapter 2.3, p. 10 for analogy.
[109] Extended Backus Nauer Form.

getOccurencyOfNthRequest,
getOccurencyOfNthRequestBackwards,
getSessionWithStartPointInInterval,
getSessionWithStartEndPointsInInterval,
getSessionWithStartEndPointsInInterval,
getSessionWithBothPointsInInterval,
getSessionWithEitherPointInInterval,
getSessionWithInteralInIntervals,
getSessionsWithLengthEqual,
getSessionsWithLengthGreater,
getSessionsWithLengthGreaterEqual,
getSessionsWithLengthLess,
getSessionsWithLengthLessEqual
getSessionsWithGreaterAndLess,
getSessionsWithGreaterEqualAndLessEqual,
getSessionsWithGreaterEqualAndLess,
getSessionsWithGreaterAndLessEqual
countSessionLength,
countSessions,
getFrequencyOfReferences,
getSessionsWithNthRequestObjectInRequestObjectSet,
getSessionsWithNthRequestObjectInRequestObjectSetBackwards,
getSessionWhichContainsARequestObjectFromSet,
getSessionWhichContainsNotARequestObjectFromSet,
getSessionWhichContainsRequestObjectsFromSet,
getSessionsWithMinAndMaxAccessToRequestObjectSet
getSessionWithStartpointAndEndpoint,
getSessionWithStartpointAndEndpointAndSessionLength,
getSessionWithStartpointEndpointSessionLengthAndOperator
getMarkovMatrix,
getEigenvector,
getReferenceStatistics,
getHistoryFromLogfiles,
getHistoryStatistics}

value → string | int
request_object_set → { ro | ro is request object }

These methods above are based on the formal model of chapter 9. The document model, the event model, and the relation between both, lead to the above implemented functions which describe the user behaviour in the information space. The method names were chosen to be self-explanatory.

In the following, we are going to group the methods and explain what each group does.

Class	NthRequestObjectInSession
Methods	FirstRequestObjectInSession, LastRequestObjectInSession, getOccurencyOfNthRequest, getOccurencyOfNthRequestBackwards
Description	These class retrieves the n^{th} object in a session. E.g. the 3^{rd} object of the session in figure 42 is T3.

Class	SessionIntervals
Methods	getSessionWithStartPointInInterval, getSessionWithStartEndPointsInInterval, getSessionWithStartEndPointsInInterval, getSessionWithBothPointsInInterval, getSessionWithEitherPointInInterval, getSessionWithInteralInIntervals,
Description	This class retrieves all sessions which are within a certain interval. The parameter which can be set is the interval. The interval to be set is a date interval. The methods mean the following: The start point of a session must be in a certain interval. The end point of a session must be in an interval. Start and end point must be in an interval. Either start point or end point must be in an interval.

Class	SessionLength
Methods	getSessionsWithLengthEqual, getSessionsWithLengthGreater, getSessionsWithLengthGreaterEqual, getSessionsWithLengthLess, getSessionsWithLengthLessEqual getSessionsWithGreaterAndLess, getSessionsWithGreaterEqualAndLessEqual, getSessionsWithGreaterEqualAndLess, getSessionsWithGreaterAndLessEqual countSessionLength,
Description	This class retrieves all sessions with a given length. The length of the sessions which should be retrieved, can be set. All Boolean operators, such as equal, greater, less, greater equal, less equal, etc. are supported. Furthermore, it counts the length of the sessions.

Class	SessionSubset
Methods	getSessionsWithNthRequestObjectInRequestObjectSet, getSessionsWithNthRequestObjectInRequestObjectSetBackwards, getSessionWhichContainsARequestObjectFromSet, getSessionWhichContainsNotARequestObjectFromSet, getSessionWhichContainsRequestObjectsFromSet, getSessionsWithMinAndMaxAccessToRequestObjectSet
Description	This class retrieves all session where a given node (e.g. T3 of figure 42) of a user session is in a predefined set of nodes The node is

	equivalent to the request object in the methods. The nodes represent the pages resp. documents of the hyperspace. The parameters which can be set, are: First, the node and second, the set of nodes.

Class	SessionVertex
Methods	getSessionsWithMinAndMaxAccessToRequestObjectSet getSessionWithStartpointAndEndpoint, getSessionWithStartpointAndEndpointAndSessionLength, getSessionWithStartpointEndpointSessionLength-AndOperator
Description	This class retrieves all sessions with given start points, end points, session lengths. A query could be: *Find all sessions which start with S1, end with E3 and have a session length of 5*. Another query could be: *Find all sessions which start with S5, end with E1 and have a session length greater than 3.*

Class	Markov
Methods	getMarkovMatrix, getEigenvector
Description	This class computes the Markov matrix and the eigenvector thereof.

Class	History
Methods	getHistoryFromLogfiles, getHistoryStatistics
Description	This class retrieves the history of the request entries in the web logfiles and does some basic statistics. From the history, it can be derived where the users come from.

Class	Hits
Method	getBeginAndEndOfSession
Description	This class derives the start date and the end date of a session.

Methods	getFrequencyOfReferences, getReferenceStatistics
Description	This method derives the frequency of each reference and does some statistics on the references.

Method	countSession
Description	This method counts the amount of sessions in a given time period.

10.4 Conceptual realisation

Filter description

In the following, we would like to explain what the filter's functions are. The filter applies in different steps. The function of the filter is to hide complexity of data from the user. It records events and assigns attributes to them. Basically, we have filters where the attribution functions depend on what occurs in the session. Other filters have an attribution function which depends on the request itself.

Application of filters:

The application helps us to pose conditional questions. This questions are:

a) Filter all sessions which have the start point in time interval [A, B]. (the time interval [A,B] determines the interval between time A (e.g 12th March 2002) and time B (e.g. 28th July 2002).
b) Filter all sessions which have end point in the time intervals [A,B] and [C,D].
c) Filter all sessions which have start point request object X and end point request object Y.

Answers are conditional expectation values. That is why we use Boolean expression with the attributes to define conditional probabilities.
Formally, the answers to those questions are conditional probabilities. Examples are:

a) The creation of the Markov matrix which gives the transition probability from one state into the other.
b) Statistics: Order F, G, H

In a further process, we use this attributes to restrict the questions we ask to the system in order to formulate conditional probabilities.
This context is logfile specific and can change accordingly or it can depend on the content.

The filters we use are described below:
We apply clustering to request[110] and to session[111] data.

- Statistics for request and session data
  We analyse the data against each other. The questions we ask are: How many requests does a session have? What is the average number of requests for a session? How many sessions do we have with the length n? How many references does a page have? What is the average number of references of a page? Make a list of the n-th referenced page, etc.

- Pre-clustering of request and session data
  We distinguish between different methods.

[110] See p. 118, point 3 in figure 54, Generic Filter Pipeline.
[111] See p. 118, point 7 in figure 54, Generic Filter Pipeline.

1. Manual clustering:
   We manually allocate each data to a particular class and enter them in the database field, a mark which indicates the affiliation to that class.

2. Clustering of metadata:
   We are able to choose metadata as a posteriori observables.
   What we do here is a two-step pre-clustering. The first step is a clustering of the a posteriori observable values. Because the a posteriori observables represent webpages, we cluster the latter into classes. The second step is a clustering of the metadata of the webpages into classes.

3. Pre-selection of a priori and/or a posteriori observables:
   With the pre-selection of the observables, we allege a cluster grouping with following rules: We cluster the requests which have the same pre-selected observable values into one class. Hence, requests with the same observable values cannot be in different classes.
   The subsequent clustering coarses the cluster grouping.

We describe the clustering formally:

φ is the clustering function
Set P: = {p | p is a precluster}
Set C: = {c | c is a cluster}
#C < #P
φ: P $\rightarrow$ C

- Filter Type 0
  Attribute of the load unit. In our case the load unit is the request.

- Filter Type 1
  Attribute allocation after overall statistics.
  The filter makes the following parameters available:
  - Start point
  - End point
  - Session length
  - Session check mark
  - Timestamp

  With these parameters, we are able to generate different views of the logfile data. We select the sessions with the favoured parameters and continue with clustering.

- Filter Type 2
  All filter type 0 attributes of a request and all filter type 0 attributes of a session as well as the order of the requests.

- Filter Type 3
  Arises from the classification of the attributes of type 0 which we undertake. (Time, IP, website (through indexing))

- Filter Type 4
  Classifications of sessions.

- Filter Type C
  The filter type C is a complex one because it incorporates Boolean expressions and all the markings discussed above.

- After the pre-selection and marking of the data, we continue clustering with the given cluster algorithm (single-, complete linkage).
  The purpose of clustering is to group similar objects into one class, where the classes are dissimilar from each other.

We also call this attribution 'marking'. Marking is a concept. This concept has a logical workflow. This workflow has, in turn, a formal workflow, which we use to formally represent load (i.e. user behaviour) in the system.
Additionally, we also implement a technical workflow. This one implements the logical one into an application.

Method	General method	Individual method
Statistics for requests	X	
Statistics for sessions	X	
Manual clustering		X
Clustering of metadata		X
Pre-selection of observables		X
Filter type 1	X	X
Filter type C	X	X
Clustering by algorithm	X	

Table 15 Classification of methods

Simple Example

The company *news.com* pays Yahoo! to appear at the first till fifth place for certain search words. The company wants that 20 percent of the users register at their site. There are two questions to be answered.

1. How does the company control whether the objectives are achieved?

2. If the objectives are not achieved, how can corrective actions be taken?

Procedure:

1. Request to Yahoo! How many users entered the search words.
2. How many of those went to the company's website.
3. How many did register?

→ Control: Were the objectives achieved ? YES/NO

If NO, following corrective action has to be taken:

4. Extension of the search words *or*
5. Extend ranking range (e.g. 1st – 10th place)
6. go to 1

10.5 Conclusion

The filter pipeline enables us to compute access event streams which are represented as requests in the web log files. These streams are used to derive user behaviour. In the technical foundations we gave an introduction to the web mining process. The similarities between the process described in that chapter and ours is that the pre-processing tasks[112] are the same. The separate steps *data cleaning*, *user identification* and *session creation* are also done by the generic filter pipeline. The difference between our analysis tool to the existing commercial web log analysis tools lies in the openness of the architecture. The generic filter pipeline has a flexible, open architecture with already build-in functionality of elementary algebraic and stochastic methods. Furthermore, it gives the possibility to continually enhance its functionality according to the analysis needs, such as further pattern discovery and analysis algorithms. Another difference is that the pipeline is an integral part of the controlling process. On the one hand, it is used by the spiral model for the evaluation of data in quadrant V while on the other hand, it evaluates user behaviour in the web controlling process.

[112] See chapter 2.3, p. 10.

11. Feasibility of the Web Controlling Framework

This chapter is going to evaluate the requirements introduced in chapter 7 and reason why they will solve the problems we have mentioned in the *Problem and Motivation* chapter.

11.1 Feasibility Evaluation

Assure a high quality of service from a stakeholder perspective
The Web Controlling Framework allows introduction of the view of any stakeholder who is concerned. Furthermore, the framework allows a more detailed classification for lead and target stakeholders as well as allowing the determination of the communication type, such as interpersonal, organizational and mass communication between those stakeholders.

Introduce a controlling method to assure a high quality of the information space
The high quality of information space is ensured through continuous improvement. It is essential because the information space is a dynamic medium since documents are removed and inserted on a regular basis. High quality, on the one hand, means that the service provided works well and do not interfere with each other. On contrary, the information distributed in the space must meet user's interest.

The web controlling process is the controlling method which ensures the demanded high quality because it is composed of an Inside-out and an Outside-in view. Former and parts of latter defines the structure and content of the information space and the latter also evaluates its usage. The process is a cycle. Every cycle improves the information space because both views interact together in order to increase high quality.

Must support the corporate objectives and user's information needs
Our interview partners state that it is important to consider the information needs of the users. Either they were asked regarding the content of the information space, or they were demanded to put appropriate information on it.[113] Moreover, the interviewees stated that it is also important to inform the employees about the company's activities. By doing so, they get more involved in the company. As a result, their motivations and satisfactions increase, which in turn, have positive effects on work efficiency and quality.[114]

Our web controlling framework supports the above asked requirements. This is done by starting the controlling process with the corporate objectives and the information needs.

[113] See [KeRi05].
[114] See [KeRi05].

Must be based on user behaviour which is derived from the log files

We stated that our approach is user oriented. This implies that user behaviour must be elicited. This is done with the generic filter pipeline.[115] It is an open, flexible architecture with build-in functionalities and the ability to expand it with an additional functionality. The pipeline processes web log files and determines the user behaviour from it.

The attainment of goals is based on continuous improvements

It is not enough to implement the information space and think that the goals will be achieved. Since the information space is dynamic and the users vary in their behaviour, a mechanism is needed to close the gap between information offer and information request. The overall goals are twofold: The Inside-out view wants to inform the employees about the company's goal and the Outside-in view want to satisfy the users' information needs. The gap can not be closed with one cycle. In fact, several cycles are necessary. With each cycle, the attained goals are converged.

The Web Controlling Framework[116] as well as the Spiral Model[117] support continuous improvement. Former connects the Inside-out and Outside-in view and allows on this level to achieve the goals by evaluating user behaviour and adapting information space accordingly. Latter backs the goals on a lower level by analysing parts of the information space and adapting it accordingly.

[115] See chapter 10, p. 111 et seqq.
[116] See chapter 7, p. 66 et seqq.
[117] See chapter 8, p. 76 et seqq.

12. Conclusions

This chapter summarizes the thesis and presents the conclusion of this work. As an outlook, we suggest topics for further research in the area of web controlling.

12.1 Contribution

The starting point of this thesis was a project conducted at a large bank in the Swiss market place. The project addressed the topics of *large Intranets*, especially the problems which occur during its use and growth. The project was conducted because of the increasingly inefficiency in its use. The outcomes of our research showed that the inefficiency was caused few reasons: Many sources of information the bank supported, neglect of the information quality, and the insufficient quality of service provided. A further problem arose from a merger of banks, in which this bank was involved. It was the heterogeneous culture of information which existed in the bank.

From this research, the topic for the thesis arose. We aim to describe a holistic approach which addresses the above mentioned problems and puts at the same time, the user at the centre of the activity recommendations.

We defined web controlling as the *continuing process of defining objectives for the usage of the information space, controlling the activities, analysing the results and taking appropriate actions to reach the first defined objectives*.

We introduced the web controlling framework as the integral approach which implements web controlling. The framework is divided into two parts: the Inside-out and the Outside-in view. The Inside-out view represents the view of the firm to information offering. In doing so, the firm has first to think about the objectives it pursues with it. These objectives must be derived from the strategy that the company is following. These two parts implement the first part of web controlling, which is, "… [the definition of] *objectives for the usage of the information space…*". The next step is to design the information space so that the objectives can be achieved.

For the design of the information space, we have introduced the web perspective schema. It represents design elements of the information space. We name them perspectives. They are derived from the principles of organizational communication. We are positive about the fact that successful information offer must comply with organizational communication. We further conducted a study to gather the state-of-the-art implementations and the deficits of the Intranet as an information channel in enterprises. We interviewed 20 enterprises for the survey. We asked questions about how they use the Intranet, what is important to them, and how they have integrated it in their activities. From the insights of both the interviews and the organizational communication, we deduce the necessary perspectives. These represent the scope for design of the information space. The perspectives enable "… [to take] *appropriate actions to reach the first defined objectives*."

We stated that our approach must be user centric. This implies that it is not enough to consider only the company's view. We also need to take the employees'

needs into account. Therefore, we introduced the Outside-in view which represents the information needs of the employees. In the upper area of the view, there is the explicit representation, and at the lower area, there is the implicit representation of the information needs. The explicit representation describes the needs that the employees explicitly demand. The implicit representation describes the information needs that the employees are looking for in the information space. This means that the employees navigate in the information space to find the necessary information. Therefore, we need to evaluate this behaviour in order to extract their needs. The question then arises how to determine user behaviour. We argue that a picture of it is stored in the web logfiles.

Furthermore, we introduced the web controlling process which integrates the Inside-out and Outside-in views and thus, ensures "*the continuing process of…*" information quality improvement.

We introduced the spiral model which defines the evaluation process of employees' information search in the information space. The spiral model is a cyclical process with five quadrants. The first quadrant defines the goals which one wants to achieve with this evaluation and introduces the method to measure the successful achievement of the goals. The second quadrant plans the actions which must be taken to achieve the goals. It shows how the information space must be designed in order to comply with the goals. Quadrant III implements the actions and measurement methods. In quadrant IV, data are collected from the information space. These are the web logfiles which are generated. Quadrant V evaluated the data and compares them with the goals. If the results agree with the goals, no further actions must be taken. If they do not agree, then the spiral model allows passing through again the cycle as described above. This evaluation process allows control and continuous improvement of the information space and complies with our definition of web controlling "*…controlling the activities…*".

We look closer to quadrant III and V, quadrant III implements the measurement methods and quadrant V evaluates the web logfiles. For the evaluation of the logfiles, we introduced the generic filter pipeline. It is an open architecture which derives user behaviour from the logfiles. This is done which elementary algebraic and stochastic methods. The architecture is open because additional functionality can be added to the filter pipeline for further processing. The functionality can be enhanced according the analysis needs and is not limited to the build in functions.

These additional functions are derived from quadrant III. Latter determines which functionality is necessary for the analysis of the data. Their concrete implementation is based on the design of the information space. Therefore, we introduced the formal models, properties and categories of the information space. Latter give a formal description of the information space, such as the document model, event model, and their relation. Based of these formal models, additional functionalities can be implemented and plugged in the generic filter pipeline. Latter enables "…[the analysis of] *the results…*" as demanded from the web controlling definition.

12.2 Outlook and further Research

12.2.1 Ongoing Transformation in Practice

The Intranet is established in the enterprises as a useful instrument. In our opinion the development of the Intranet will operate according to the transformation phases in which we have described: Information channel, communication channel, work instrument, knowledge management tool. Firms use it as an information channel, but with time they tend to enhance its functionality. The reason is that they see the Intranet as a valuable opportunity to improve the organizational effectiveness. From a pure information channel, it will become a communication channel. Moreover, latter enables two-way communication and collaborative work.

This transformation will take several years. We are convinced that the process will not stop after the last phase, but will go on further. At this point, we are not able to identify the next transformation phases of the information space. The identification of the next phases must be left to the further researches.

12.2.2 Closing the Cap

The implementation of web controlling framework and of the spiral model faces great difficulties in their actual usage. On the one hand, it is the gap between the strategic goals and web controlling process and on the other hand, it is between the spiral model and the technical evaluation. The closure of these gaps should be the purpose of future researches. In recent years, we have originally drafted so-called patterns for the OO-development and achieved all the levels of the system design. Examples of these are architecture pattern, integration pattern, and security pattern.

Patterns are used to facilitate the communication between experts and for education. With a certain delay also anti-patterns have been introduced, e.g. by Brown et al. ([BrMa98]), where they have brought about such as anti-pattern for versioning management, and etc.

Hence, the question is whether there are patterns and anti-patterns for the local design of information spaces, especially for the navigation structure, and whether these are re-produced in the log files and if they can be identified there.

An important application example for anti-pattern is the elimination of the apparent and inevitable design flaw in its use.[118]

Riedl ([Riedl01c]) shows how path information and information about web page content can be combined in order to identify user behaviour in hyperspaces. Riedl shows how the semantic relation mapping analysis is produced. The question is to what extent the closure of these gaps on a technical aspect would be applied and how much of the spiral model could be implemented. Also, some other questions are whether potential patterns and the semantic notations from these connections could be identified. In doing so, we are also able to offer the application of ontology.

[118] See [Ried01b].

Further research must be conducted in the evaluation of the web perspective schema[119]. This empirical study has the purpose of validating the various perspectives in cooperation with some enterprises. It must be checked out if the latter are valid for all types of enterprises. After having reviewed the perspectives, a standard library could be derived from these according to ITIL (IT Infrastructure Library).

12.2.3 Considering Culture

The Paper on role structures ([Ried99b]) was intended to show that the theory on social role structure was interpreted in the context of the users' perspective in the field of information spaces. The idea which was developed in chapter 10 on Generic Filter Pipeline is to calculate the classes which are based on the social role structure.

A merger between two enterprises means that there also exist two different corporate cultures, in which the user behavior should eventually integrate. Hence, there are two questions. The first question is whether it is possible to follow the incorporation of the cultures with the help of distance for half groups ([Boga75]). Next, it is the question of whether it is also possible to coherently join the Web Controlling Framework and the Spiral Model to develop an information provisioning culture for controlling purposes.

From basic approaches from Rühli[120], culture is regarded as one of the most important aspects. In 4+2 approach of Joyce et al. ([JoNo04]), culture is rated as one of the four primary formation factors and from Riedl[121], the cultural aspect is clearly emphasized in knowledge management. Thus, as a control for culture development, especially in dealing with information and knowledge, it would especially be valuable tool in business management. So far, generic measurement methods regarding this area are missing or are too general in respect of the intellectual capital.
Our approach has the advantage that it is based on user behavior in the Intranet, which is available virtually in all large enterprises. For this reason, there were in fact no absolute measures produced, yet the methods were independent and easily applicable between departments within the firm. Furthermore, it could also incorporate the above mentioned part about the quality standards.

A greater importance for the enterprise's future is the question whether we can control corporate's information democratization[122] with our spiral model and with our web controlling framework. Marchand et al. ([MaKe02]) consider this democratization as an essential goal of a technical information system.

[119] See chapter 7.3, p. 70.
[120] See [Rühl96].
[121] See [Ried02].
[122] See chapter 6.3, p. 64.

Appendix

A Link graphs of several websites

Website of the Institute for Information Technology, University of Zurich, Database Technology Group
http://www.ifi.unizh.ch/groups/dbtg/

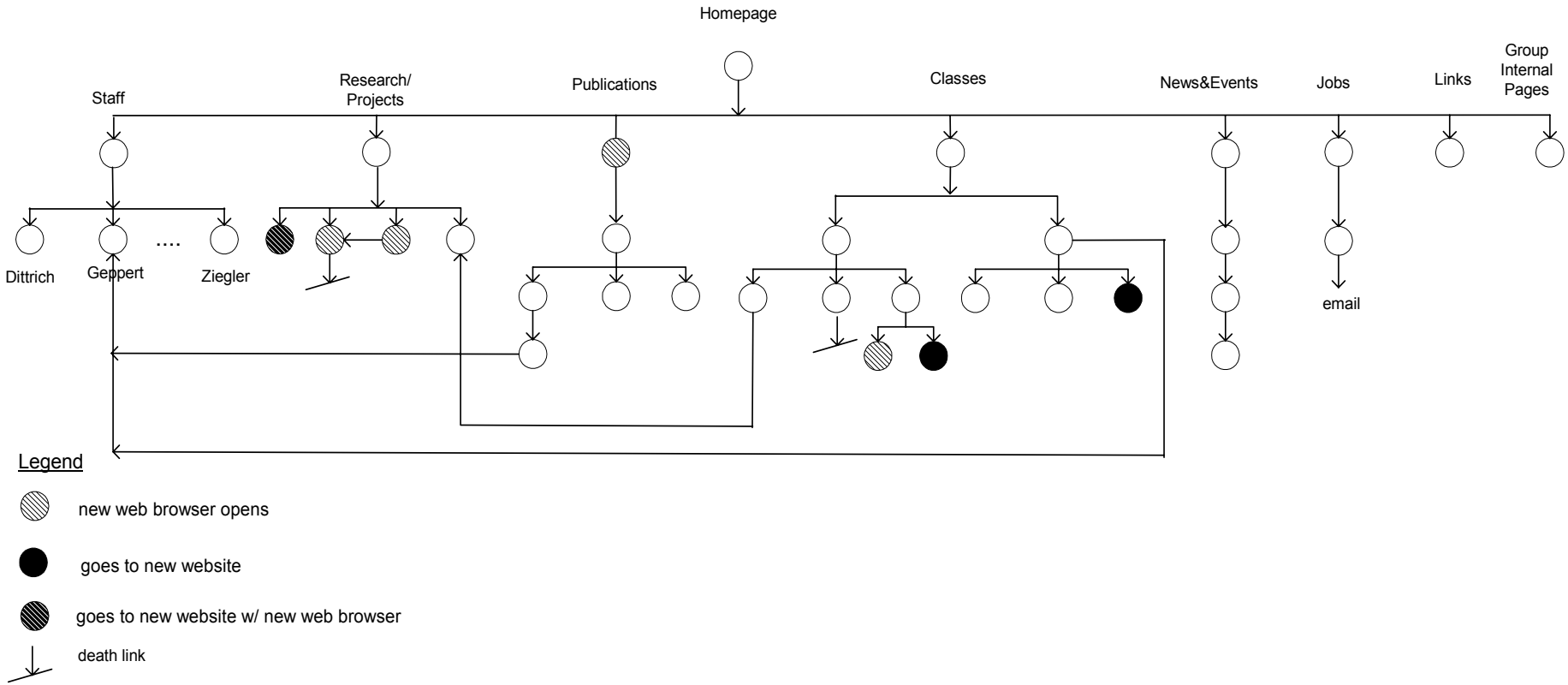


Figure 55 *Website of the Institute for Information Technology*

Commercial website of the company Luchsinger Mathematics
http://www.luchsinger-mathematics.ch

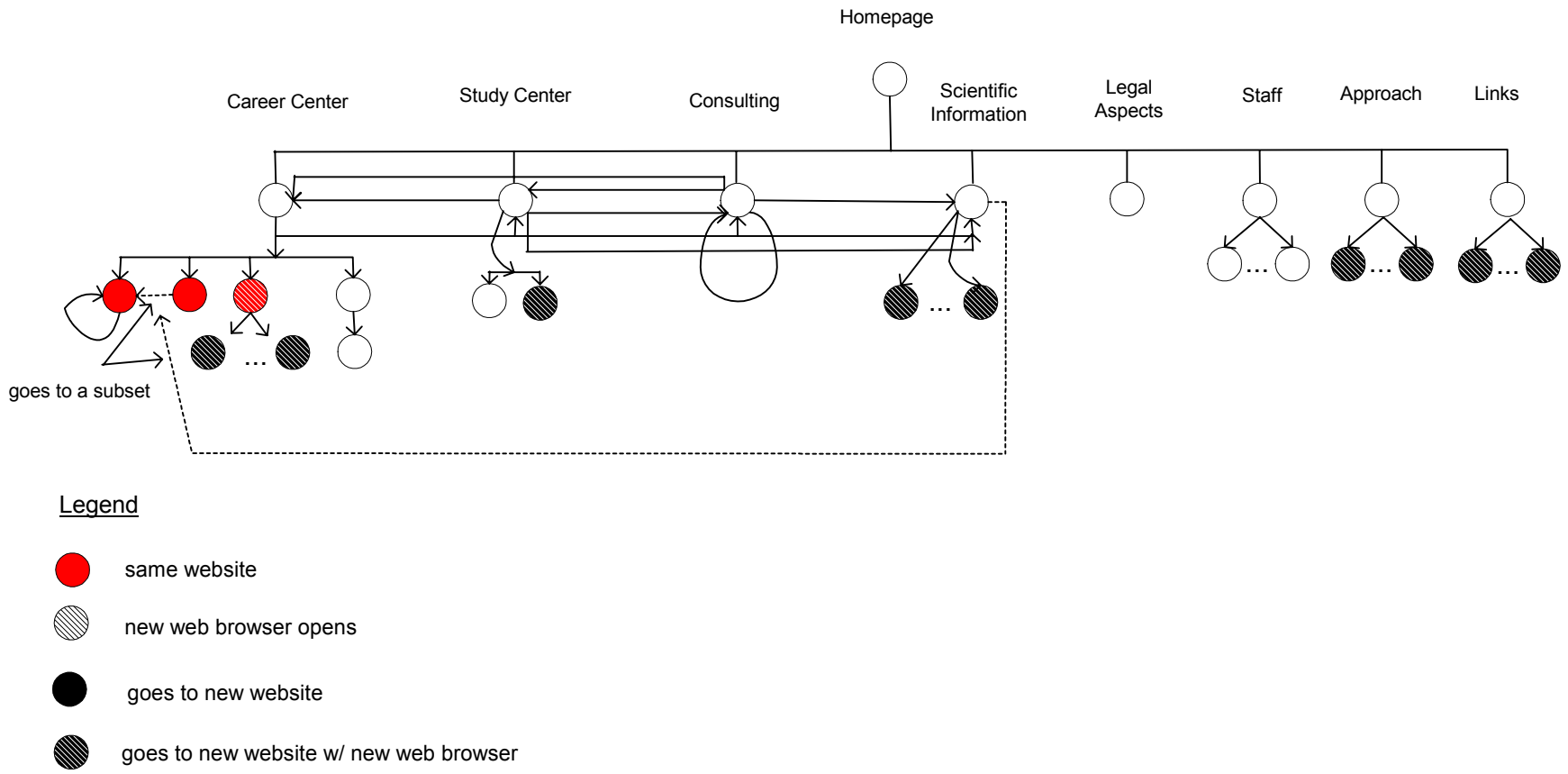


Figure 56 W*ebsite of Luchsinger Mathematics*

Website of Zurich's Cynical Theatre Guide
http://www.ifi.unizh.ch/groups/richter/people/riedl/theater.html

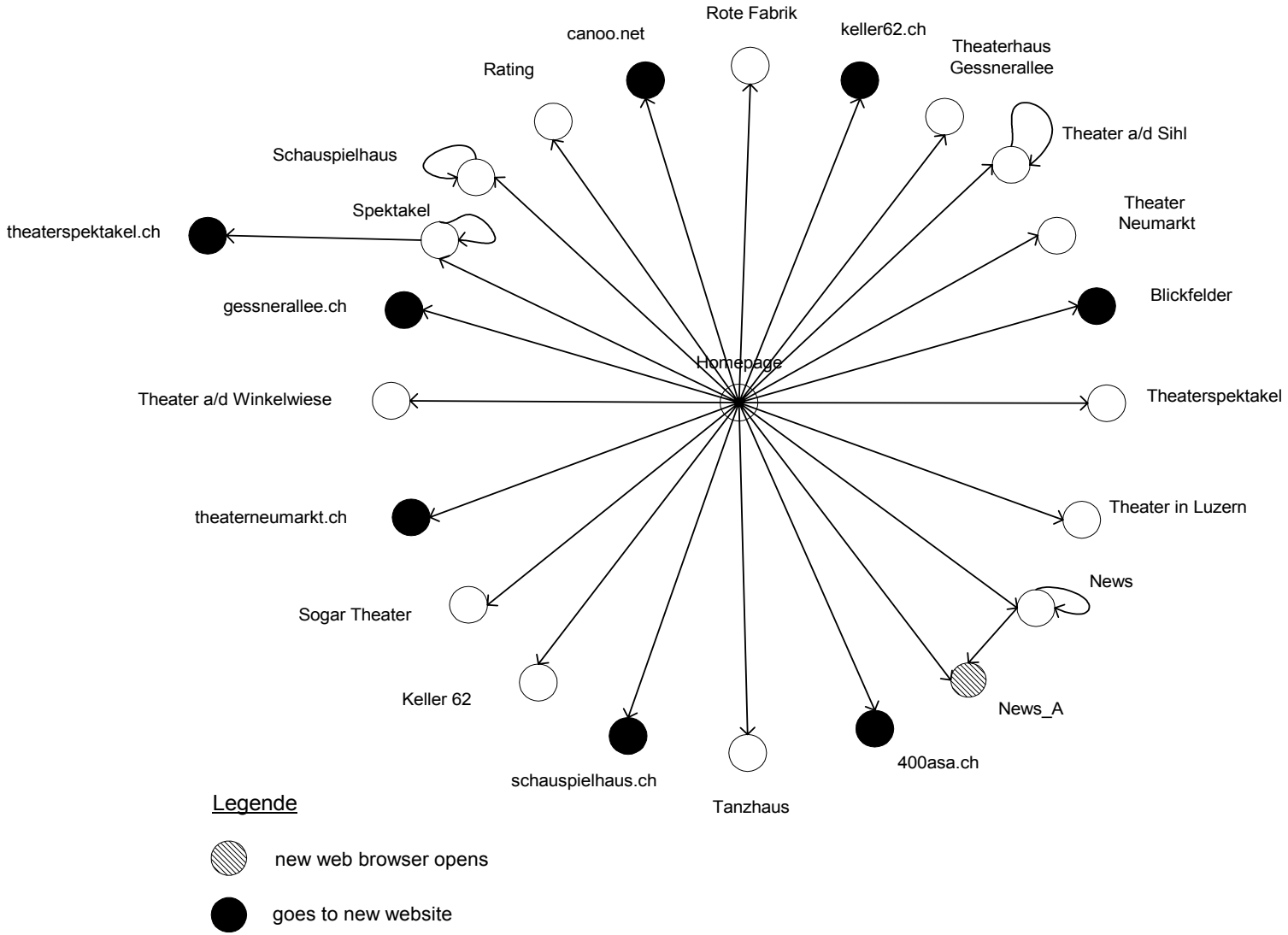


Figure 57 *Website of Zurich's Cynical Theatre Guide*

Website of the European project FASME
https://members.fasme.org/

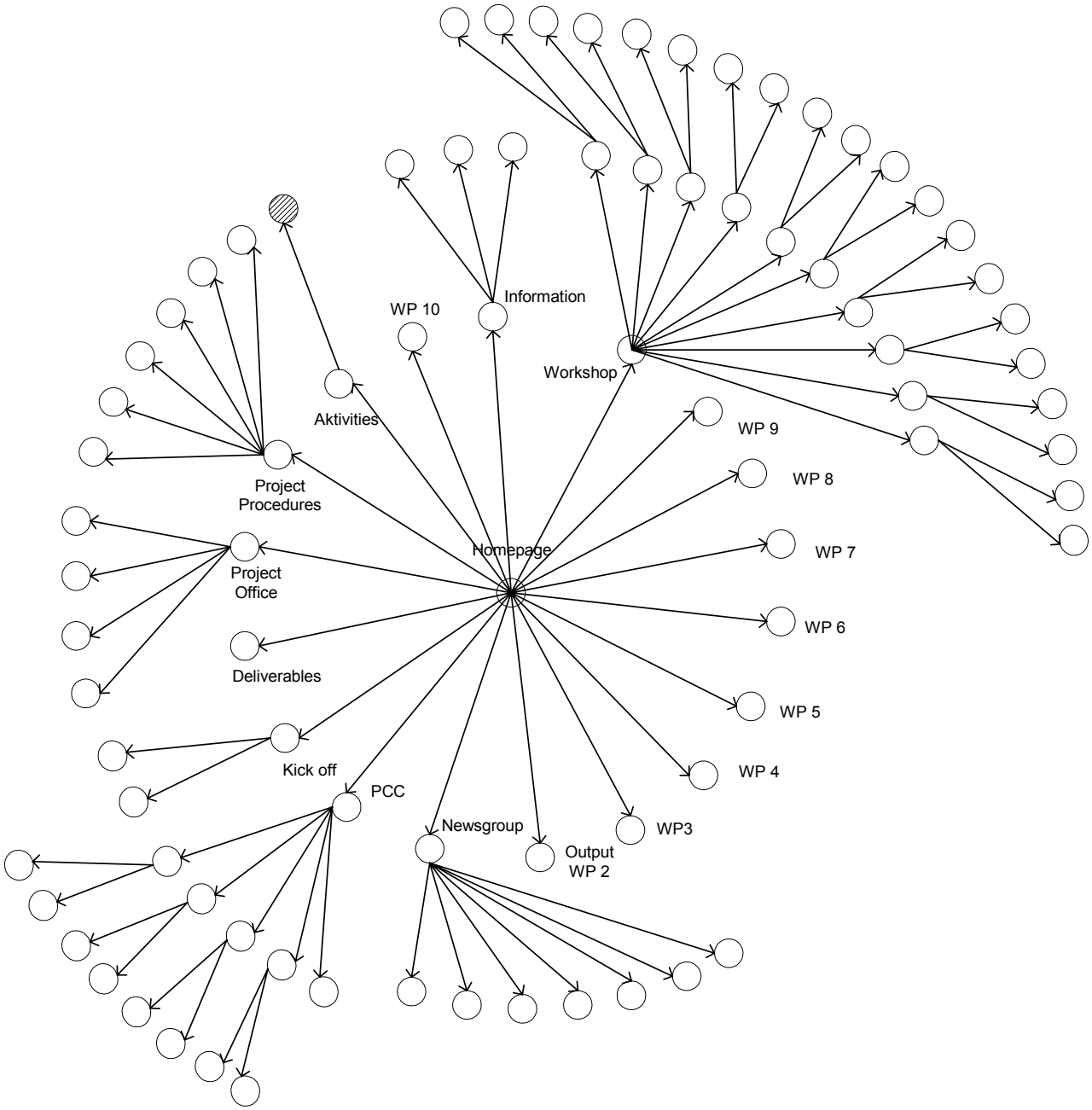


Figure 58 Website of the EU Project "FASME"

Legende

new web browser opens

B Interview Questions

The full interviews are given in the technical report ([KeRi05]). Below are the questions we have posed to the interview partner.

Questions

1. *What function does the Intranet have in your company?*

2. *For which purpose will you pursue with the Intranet?*

3. *How are the corporate goals integrated into the Intranet management resp. what corporate goals does the company pursue?*

4. *Did the organizational structures resp. processes change due to the usage of the Intranet?*

5. *Did the organizational information and communication structures resp. processes change due to the usage of the Intranet?*

6. *Did the labor organization resp. processes change due to the usage of the Intranet?*

7. *Does the company conduct trainings?*

8. *What are the topics and contents of the Intranet?*

9. *According to what criteria do you design the Intranet?*

10. *What tools do you use and what steps do you take in order to ensure the quality of the Intranet? I.e. correct and up to date content, removal of old or wrong information etc.*
    If yes, which?

11. *Do you think that the Intranet is of any benefit for your company?*
    If yes, which?

12. *Do you think that the employees benefit from the Intranet?*
    If yes, in what way?

13. *How do you measure the benefit from the Intranet for your company?*

14. *Do you know how the Intranet is used by the employees? Did you implement a controlling mechanism?*

15. *Do you get any feedback from the employees concerning navigation, content and usability of the Intranet?*

16. *Do you have implemented feedback mechanisms resp. how do you get feedback from the employees?*

17. *Do you involve the employees in the design of the Intranet resp. in parts of it? How do you involve them?*

18. *Do you conduct re-designs of the Intranet?*
    If yes, at whose suggestion?

19. *Which relevance has (resp. will have) the Intranet for your company in your opinion?*

20. *Where are the strengths and weaknesses of the Intranet as a communication media?*

C Abbreviations

BOS	Begin of Session
CMS	Content Management System
DB	Database
DMS	Document Management System
doc	Microsoft Document Datatype
EOS	End of Session
et seq.	and the following
et seqq.	and the following
FASME	*F*acilitating *A*dministrative *S*ervices for *M*obile *E*uropean
FTP	File Transfer Protocol
GIF	Graphics Interchange Format
HTML	Hypertext Mark-Up Language
HTTP	Hypertext Transfer Protocol
IFI	Department of Informatics
IP	Internet Protocol
IT	Information Technology
JPEG	Joint Photographic Experts Group
Link	Hyperlink
LOS	Length of Session
MPEG	Moving Picture Experts Group
OO	Object Oriented
PDF	Portable Document Format
pdf	Portable Document Format
PM	Project Manager
PS	PostScript
ps	PostScript
ROI	Return On Investment
s.	see
Site	Website
SMTP	Simple-Mail-Transfer Protocol
TCP	Transmission Control Protocol
vs.	versus
WWW	World Wide Web
ZZT	Zürichs Zynischer Theater-Index

D Glossary

Capability Model	Model which defines the special access rights between subjects and objects according to [Rich85].
Document	A document is a HTML page, a PDF file, a multimedia file or an immaterial product (e.g. hotel reservation). A document contains an URL and can be reached through this URL.
Extranet	"An Extranet is a private network that uses the Internet protocols and the public telecommunication system to share a business's information, data or operations with external suppliers, vendors or customers. An Extranet can be viewed as the external part of a company's Intranet." Definition by www.easynet.com/investorinfo/investorinfo_glossary.asp "An Extranet is a collaborative network that uses Internet technology to link businesses with their suppliers, customers, or other businesses that share common goals. An Extranet can be viewed as part of a company's Intranet that is made accessible to other companies or that is a collaboration with other companies. The shared information might be accessible only to the collaborating parties or, in some cases, might be public." definition by www.course.com/careers/glossary/internet.cfm
Holschuld	Definition from http://dict.leo.org: "debt to be discharged at the domicile of the debtor". In our context Holschuld means that the employee resp. the user is responsible and obliged to get the necessary information.
Hyperspace dynamic	The set of all documents which are interlinked with each other as well as all documents generated by search engines.
Hyperspace static	The set of all documents which are interlinked with each other.
Intranet	"A private network inside a company or organization, which uses software like that used on the Internet, but is for internal use only, and is not accessible to the public. Companies use Intranets to manage projects, provide employee information, distribute data and information, etc." definition by www.getnetwise.org/glossary.php "A private network inside a company or organization that uses the same kinds of software that you would find on the public Internet, but that is for internal use only. Firewalls keep unauthorized Internet traffic off an Intranet. As the Internet has become more popular, many of the tools used on the Internet are being used in private networks. For example, many companies have web servers that are available only to employees. Note that an Intranet may not actually be an Internet, it may simply be a network." Definition by

	www.gtscompanies.com/glosscomp.html "A Web site that is hosted on a local network and is, therefore, not accessible by anyone outside the network. Intranets work just like public Web sites. Offices can include any text, graphics, and hyperlinks it chooses. Intranets can also provide access to external Web sites, even though external users cannot access the Intranet." Definition by www.congressonlineproject.org/glossary.html

E Bibliography

[AdTu01] Adomavicius, G., Tuzhilin, A., *Using Data Mining Methods to Build Customer Profiles*, IEEE Computer, vol. 34 num. 2 (2001) 74-82.

[Ande83] Anders, W., *Kommunikationstechnik und Organisation: Perspektiven für die Entwicklung der organisatorischen Kommunikation*, CW-Publikationen, München, 1983.

[Armb92] Armbrecht, W., *Innerbetriebliche Public Relations: Grundlagen eines situativen Gestaltungskonzepts*, Westdeutscher Verlag, Opladen, 1992.

[Baac73] Baacke, D., *Kommunikation und Kompetenz*, Juventa Verlag, München, 1973.

[Bart69] Bartram, P., *Die innerbetriebliche Kommunikation*, Erich Schmidt Verlag, Berlin, 1969.

[Boeh86] Boehm, B.W., *A spiral model of software development and enhancement*, ACM Sigsoft Software Engineering Notes, 11, 4, 1986.

[Boga75] Bogart, K.P., *Preference Structures: Distances between Asymmetric Relations*, SIAM J Appl. Math 29, 1975.

[BrMa98] Brown, W.J, Maveau, R.C., Hays W., McCormick, S., Mowbray, T.J., *AntiPatterns: Refactoring Software, Architectures, and Projects in Crisis*, Wiley; 1st edition, 1998.

[Bruh95] Bruhn, M., *Integrierte Unternehmenskommunikation: Ansatzpunkte für eine strategische und operative Umsetzung integrierter Kommunikationsarbeit*, Schäffer-Poeschel Verlag, Stuttgart, 1995.

[Bruh97] Bruhn, M., *Kommunikationspolitik*, Verlag Franz Vahlen, München, 1997.

[BuHö92] Burkhart, R., Hömberg, W.(Hg.), *Kommunikationstheorien: Ein Textbuch zur Einführung*, Wilhelm Braunmüller, Wien, 1992.

[BüMu98] Büchner, A. G., Mulvenna, M. D., *Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining*, SIGMOD Record, 27(4):54—61, 1998.

[CoMo97] Cooley, R., Mobasher, B., Srivastava, J., *Web Mining: Information and Pattern Discovery on the World Wide Web*, In Proceedings of the Ninth IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), November 1997.

[CoMo99] Cooley R., Mobasher B., Srivastava J., *Data Preparation for Mining World Wide Web Browsing Patterns*, Journal of Knowledge and Information Systems, 1(1), 1999.

[DaLe86] Daft, R.L., Lengel, R.H., *Organizational information requirements, Media Richness and structural design*, in: Management Science, Vol. 32, No. 5, May 1986, p. 554-571.

[DePo94] DeSanctis, G., Poole, M. S., *Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory*, Organization Science, Vol. 5, No. 2, May 1994, p. 121-148.

[DeVa99] Dennis, A. R., Valacich, J.S., *Rethinking Media Richness: Towards a Theory of Media Synchronicity*, in: Proceedings of the 32nd Annual Hawaii International Conference on System Sciences, 1999, IEEE Computer Society, Los Alamitos, 10p.

[DyMe02] Dyllick, T., Meyer, A., *Kommunikationsmanagement*, in: Dubs, R., Euler, D., Rüegg-Stürm, J., *Einführung in die Managementlehre*, Verlag Paul Haupt, Bern, 2002.

[EdMa97] Edvinsson, L., Malone, M.S., *Intellectual Capital: Realizing Your Company's True Value by Finding Its Hidden Brainpower*, Harper Business, 1997.

[EiVa03] Eirinaki, M., Vazirgiannis, M., *Web Mining for Web Personalization*, ACM Transcations on Internet Technology, Vol. 3, No. 1, February 2003, Pages 1-27.

[EuRo87] Euske, N.A., Roberts, K.H., *Evolving Perspectives in Organizational Theory: Communication Implications*, in: Jablin, F.M., Putnam, L.L., Roberts, K.H., Porter L.W., *Handbook of Organizational Communication*, Sage Publications, 1987.

[EvWu00] Evans, P., Wurster, T.S., *Blown to bits: how the new economics of information transforms strategy*, Harvard Business School Press, Boston, 2000.

[GaSc02a] Gaul, W., Schmidt-Thieme, L., *Recommender Systems based on User Navigational Behavior in the Internet*, Behaviormetrika, Vol. 29, No.1, 2002, p. 1-22.

[GaSc02b] Gaul, W., Schmidt-Thieme, L., *Web Controlling and Recommendersysteme*, In Hippner, H., Merzenich, M., Wilde, K. D. (Hrsg.), Handbuch Web Mining im Marketing, Vieweg, 1. Auflage, 2002, S. 235-247,.

[Gent02] Gentsch, P., *Personalisierung der Kundenbeziehung im Internet – Methoden und Technologien*, In Hippner, H., Merzenich, M., Wilde, K. D. (Hrsg.), Handbuch Web Mining im Marketing, Vieweg, 1. Auflage, S. 267-307, 2002.

[GrHe95] Greenbaum, H.H., Hellweg, S.A., Falcione, R.L., *Organizational Communication Evaluation: An Overview, 1951-1981*, in: Goldhaber, G.M., Barnett, G.A., *Handbook of Organizational Communication*, Ablex, New Jersey, 1995.

[GrMi00] Gryza, C., Michaelis, T., Walz, H., *Strategisches Informationsmanagement, Das Intranet als entscheidender Faktor in der Unternehmenskommunikation*, Publicis MCD Verlag, München, 2000.

[Grun92] Grunig, J.E., *Excellence in Public Relations and Communication Management*, Lawrence Erlbaum Associates, Publishers, 1992.

[Gysi02] Gysi, R., *Auswertung von Logfiles in Suchmaschinen und Webservern*, Semesterarbeit, Institut für Informatik der Universität Zürich, 2002.

[Hast89] Hastings, K.J., *Introduction to the mathematics of operations research*, Marcel Dekker Inc., 1989.

[Hein01] Heine, M., *E-Commerce-Konzepte und Web-Controlling – durch Kundenorientierung zum erfolgreichen E-Commerce-Controlling*, In Hermanns, A., Sauters M. (Hrsg), Management-Handbuch Electronic Commerce, 2. Auflage, 2001, S. 159-171.

[HiMe02] Hippner, H., Merzenich, M., Wilde, K. D., *Grundlagen des Web Mining,* In Hippner, H., Merzenich, M., Wilde, K. D. (Hrsg.), Handbuch Web Mining im Marketing, Vieweg, 1. Auflage, S. 4-28, 2002.

[Hoff01] Hoffmann, C., *Das Intranet – Ein Medium der Mitarbeiterkommunikation*, UVK Verlagsgesellschaft, Konstanz, 2001.

[HuSä02] Huber, K.-P., Säuberlich, F., Böhm, C., *Kennzahlenbasiertes Web Controlling mit einer Web Scorecard*, In Hippner, H., Merzenich, M., Wilde, K. D. (Hrsg.), Handbuch Web Mining im Marketing, Vieweg, 1. Auflage, 2002. S. 407-419.

[JoNo04] Joyce, W., Nohria, N., Roberson, B., *What Really Works: The 4+2 Formula for Sustained Business Success*, HarperBusiness, 2004.

[KaNo96] Kaplan, R.S., Norton, D.P., *The Balanced Scorecard: Translating Strategy into Action*, Harvard Business School Press; 1996.

[KeRi05] Kefos, Ch., Riedl, R., *Interviews about the Practical Implementation of the Intranet in Enterprises*, Technical Report, University of Zurich, 2005.

[KrZe98] Krzeminski, M., Zerfass, A., *Interaktive Unternehmenskommunikation*, F.A.Z.-Institut, Frankfurt am Main, 1998.

[KüMü98] Kübler, B., Müller, Th., Winzenried, R., *Seifenblasen oder Magische Medien? Internet und Intranet bei der Deutschen Shell Aktiengesellschaft*, in: Krzeminski, M., Zerfass, A., *Interaktive Unternehmenskommunikation*, F.A.Z.-Institut, Frankfurt am Main, 1998.

[KuWo00] Kuppinger, M., Woywode, M., *Vom Intranet zum Knowledge Management: Die Veränderung der Informationskultur in Organisationen*, Hanser, 2000.

[Leib03] Leibrandt, G., *E-Interview „Web-Controlling – Transparente Erfolgsmessung von Online-Aktivitäten“*, http://www.competence-site.de/ebusiness.nsf/85ba6c60fa84ab0bc1256a8c003146c9/5cdb4277186596a5c1256d480048efca, 2003.

[Luhm85] Luhmann, N., *Soziale Systeme: Grundriss einer allgemeinen Theorie*, Suhrkamp, Frankfurt am Main, 1984.

[LuRi00] Lueg, Ch, Riedl, R., *How Information Technology Could Benefit from Modern Approaches to Knowledge Management*, Proceedings of the 3rd Conference on Practical Applications of Knowledge Management - PAKM 2000, Basel 2000

[MaBe01] Mayer, R., Bensberg, F., Hukemann, A., *Web Log Mining als Controlling-Instrument für Online-Shops*, In Controlling, Verlag Valen, München, 3/ 2001. S. 156-166.

[Mach90] Macharzina, K, *Informationspolitik: Unternehmenskommunikation als Instrument erfolgreicher Führung*, Gabler, Wiesbaden, 1990.

[MaKe02] Marchand, D.A., Kettinger, W.J., Rollins, J.D., *Information Orientation: The Link to Business Performance*, Oxford University Press, 2002.

[Mast02] Mast, C., *Unternehmenskommunikation: Ein Leitfaden*, Lucius & Lucius, Stuttgart, 2002.

[Mato00] Matousek Jan M., *Werkzeuge zum Internet-Profiling*, Diplomarbeit, Institut für Informatik der Universität Zürich, 2000.

[McWi93] McQuail, D., Windahl, S., *Communication models for the study of mass communications*, 2nd edition, Longman, London and New York, 1993.

[Meie02] Meier, Ch., *Kommunikation in räumlich verteilten Teams: Videoconferencing bei Technics*, in: Thimm, C. (Hrsg.), *Unternehmenskommunikation offline/online*, Lang, Frankfurt am Main, 2002.

[MeWe01] Meyer, M., Weingärtner, S., Jahke, T., Lieven, O., *Web Mining and Personalisierung in Echtzeit*, Arbeitspapier, Seminar für empirische Forschung und Unternehmensplanung, 5/ 2001.

[MoCo00] Mobasher, B., Cooley, R., Srivastava, J., *Automatic Personalization Based on Web Usage Mining*, Communications of the ACM, 43(8):142—151, 2000.

[Mong87] Monge, P.R., *The Network Level of Analysis*, in: Berger, C.R., Chaffee, S.H., *Handbook of Organizational Science*, London, 1987, p. 241 et seq.

[MüLe03] Müller-Stewens, G., Lechner, Ch., *Strategisches Management – Wie strategische Initiativen zum Wandel führen*, Schäffer Poeschel Verlag Stuttgart, 2. Auflage, 2003.

[NaKa00] Nakayama, T. Kato, H., Yamane, Y. *Discovering the gap between Web site designers' expectations and users' behavior*, Computer Networks 33, 2000, p. 811-822

[NuGu98] Nusch, F., Guntrum, N.L., *Hoechst Online Relations – Digitalisierung der Konzernkommunikation und veränderte interne Kommunikation*, in: Krzeminski, M., Zerfass, A., *Interaktive Unternehmenskommunikation*, F.A.Z.-Institut, Frankfurt am Main, 1998.

[PiRe87] Picot, A., Reichwald, R., *Bürokommunikation: Leitsätze für den Anwender*, AIT Angewandte InformationsTechnik Verlag, Hallbergmoos, 1987.

[Rich85] Richter, L., *Betriebssysteme*, B.G. Teubner Stuttgart, 2. Auflage, 1985.

[Ried01a] Riedl, R., *Need for Trace Benchmarks*, In, Rudi Eigenmann, editor, Performance Evaluation with Realistic Applications, MIT Press 2001.

[Ried01b] Riedl, R., *Customer-Centered Models for Websites and Intranets*, Institute of Information Technology, University of Zurich, HPCN 2001.

[Ried01c] Riedl, R., *Event Mining in Virtual Markets: Market Research, Knowledge Engineering, Information Agents, and Social Role Structures*, to appear in Proc. IFIP Working Conference on E-Business, Salzburg 2001, Kluwer Publishing.

[Ried02] Riedl, R. *Some Critical Remarks in Favour of IT-Based Knowledge Management,* in: Special issue on Knowledge Management, edited by C. Lueg, Informatik/Informatique, 2002

[Ried99a] Riedl, R., *The Impact of Workload on Simulation Results for Distributed Transaction Processing*, HPCN Europe '99, Amsterdam 1999.

[Ried99b] Riedl, R. *Usage of Trace Data for the Deduction of Role Structures and Comparison of Knowledge Societies*, CT' 99, San Francisco, 1999.

[Romm02] Rommert, F.-M., *Hoffnungsträger Intranet: Charakteristika und Aufgaben eines neuen Mediums in der internen Kommunikation*, Verlag Reinhard Fischer, 2002.

[Rüeg03] Rüegg-Stürm, J., *Das neue St. Galler Management-Modell*, 2. Auflage, Verlag Paul Haupt, 2003.

[Rühl96] Rühli, E., Unternehmensführung und Unternehmenspolitik, Paul Haupt, 1996.

[SaGa03] Savaresi, S. M., Garatti, S., Bittanti, S., La Brocca, L., *Data-Mining of a Large Virtual Community: Relationships Between the Users DB and the Web-Log File*, In Daniel Barbará, Chandrika Kamath, editors, Proceedings of the Third SIAM International Conference on Data Mining, San Francisco, 2003.

[Saxe99] Saxer, U., *Organisationskommunikation aus kommunikationswissenschaftlicher Sicht: Eine Standortbestimmung,* in: Szyszka, P., *Öffentlichkeit: Diskurs zu einem Schlüsselbegriff der Organisationskommunikation*, Westdeutscher Verlag, 1999.

[Schi02] Schick, S., *Interne Unternehmenskommunikation, Strategien entwickeln, Strukturen schaffen, Prozesse steuern*, Schäffer-Poeschel Verlag, Stuttgart, 2002.

[Schw01] Schwabe, G., *Theorien zur Mediennutzung bei der Gruppenarbeit*, in: Schwabe, G., Streitz, N., Unland R. (Hrsg.), *CSCW-Kompendium: Lehr- und Handbuch zum computerunterstützten kooperativen Arbeiten*, Springer, 2001.

[ScKr96] Schwabe, G., Krcmar, H., *Needs Driven Approach: Ein Herausforderung der Telekooperation*, Proceedings der DCSCW, 1996.

[Sorg79] Sorg, H. P., *Aufgabe der innterbetrieblichen Information für Mitarbeiter und Messung ihrer Wirksamkeit*, Dissertation, Berlin, 1979.

[SrCo00] Srivastava, J., Cooley, R., Deshpande, M., Tan, P., *Web Usage Mining: Discovery and Applications of Usage Patters from Web Data*, ACM SIGKDD, Volume 1, Issue 2, page 22, Jan 2000.

[SrYa01] Srikant R., Yang, Y., *Mining Web logs to improve Website Organization*. In Proceedings of the Tenth International World Wide Web Conference, p. 430-437, Hong Kong, May 2001

[Stoh95] Stohl, C., *Organizational Communication: Connectedness in Action*, Sage Publications, 1995.

[StSc99] Stürzebecher, G., Schröter, Th., *Das Junior-Net der Bertelsmann AG: Entwicklung eines Sub-Intranets zur Förderung der internen Kommunikation für Führungsnachwuchs*, in: Jäger, W., Jäger, M. (Hrsg.), *Personalarbeit im Intranet*, Luchterhand, 1999.

[Syst02] SYSTOR, *Web-Controlling – Transparenz per Mausklick: Der Durchblick in Echtzeit*, http://www.systor.ch/know_bizpub_ebusiness_webcontrolling, 2002.

[Szys99] Szyszka, P., *Organisationskommunikation*, Instutute of Journalism and Media Research, University of Zurich, 1999.

[Tane02] Tanenbaum, A.S., *Computer Networks*, 4th edition, Prentice Hall, 2002.

[Thei03] Theis-Berglmair, A., *Organisationskommunikation: Theoretische Grundlagen und empirische Forschungen*, LIT Verlag, 2. Auflage, 2003.

[Thim02] Thimm, C. (Hrsg.), *Unternehmenskommunikation offline/online*, Lang, Frankfurt am Main, 2002.

[Tonn98] Tonnemacher, J., *Mitarbeiterkommunikation*, in: Merten, K., Zimmermann, R. (Hg.), *Das Handbuch der Unternehmenskommunikation*, Köln/Kriftel, S. 99-105, 1998.

[WaBe03] Watzlawick, P., Beavin, J.H., Jackson, D.D., *Menschliche Kommunikation: Formen, Störungen, Paradoxien*, 10. unveränderte Auflage, Verlag Hans Huber, 2003.

[Weve96] Wever, U.A., *Interne Erfolgsbereitschaft als externer Erfolgsfaktor*, in:

Bentele, G., Steinmann, H., Zerfass, A. (Hrsg.), *Dialogorientierte Unternehmenskommunikation: Grundlagen – Praxiserfahrungen – Perspektiven*, VISTAS Verlag, 1996.

[Zerf96a] Zerfass, A., *Unternehmensführung und Öffentlichkeitsarbeit*, Opladen, Westdeutscher Verlag, 1996.

[Zerf96b] Zerfass, A., *Dialogkommunikation und strategische Unternehmensführung*, in: Bentele, G., Steinmann, H., Zerfass, A. (Hrsg.), *Dialogorientierte Unternehmenskommunikation: Grundlagen – Praxiserfahrungen – Perspektiven*, VISTAS Verlag, 1996.

Curriculum Vitae

Personnel data

Name:	Christos Kefos
Date of Birth:	31. August 1968
Place of Birth:	St. Gallen, Switzerland
Home town:	Widnau (St. Gallen) and Greece

Education

2001-2005	External PhD Student at the University of Zurich Department of Informatics
1998-2001	Research Assistant at the University of Zurich Department of Informatics Research Assistant to Prof. Dr. L. Richter System Architecture & Software Group
1987-1993	Swiss Federal Institute of Technology Zurich (ETH) Studies in Electrical Engineering
1983-1987	Gymnasium Heerbrugg, Heerbrugg (SG)
1975-1983	Primary and Secondary School, Widnau (SG)

Working experience

2001-2002	Cap Gemini Ernst & Young, Switzerland Project Manager and IT Architect
1997-1998	Cambrigde Technology Partners, Zurich Senior Consultant for Business Information Systems
1995-1997	ISMC AG , Zurich Software Engineer